STIC-Biotech/ChemLib 169433 MEJ

From: Chan, Christina
Sent: Monday, October 24, 2005 5:39 PM
To: Walicka, Malgorzata; STIC-Biotech/ChemLib
Subject: RE:

Please rush. Thanks

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
(571)-272-0841
Remsen, 3E89

-----Original Message-----
From: Walicka, Malgorzata
Sent: Monday, October 24, 2005 3:50 PM
To: Chan, Christina
Subject:

Christina, please authorize rush search of 3 short popeptides of SEQ ID NO:2, 3, and 9 in applicaiton 09/774,954. Please also against DNA data bases. Thank you.

Malgorzata

Malgorzata A. Walicka, Ph.D.
Patent Examiner
Art Unit 1652, Recombinant Enzymes
USPTO, Remsen Building, Room 2C76
400 Dulany St.
Alexandria, VA 22313
Mail Room 2C70
Tel. (571) 272-0944, fax (571) 273-0944

2 aa 365
3 / 61
9 aa 61

MEJ

****
Searcher:_____
Searcher Phone:_____
Date Searcher Picked up:_____
Date completed:_____
Searcher Prep Time:_____
Online Time:_____

****
Type of Search
NA#_____ AA#:_____
S/L:_____ Oligomer:_____
Encode/Transl:_____
Structure #:_____Text:____
Inventor:_____ Litigation:___

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM:_____
WWW/Internet:_____
Other (Specify):_____

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:21:23 ; Search time 210.606 Seconds
(without alignments)
887.482 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGYLLLYCPCMGR.....QGRPSSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1979	100.0	388	1	OFU1_HUMAN	Q9h488 homo sapien
2	1979	100.0	388	1	OFU1_PANTR	Q6ev69 pan troglod
3	1895	95.8	391	2	Q7YRE7	Q7yre7 bos taurus
4	1878	94.9	391	2	Q6EV66	Q6ev66 bos taurus
5	1854	93.7	395	2	Q6EV70	Q6ev70 rattus norv
6	1853	93.6	378	2	Q7YRZ3	Q7yrz3 sus scrofa
7	1832	92.6	393	1	OFU1_MOUSE	Q91zw2 mus musculu
8	1829	92.4	393	2	Q8C8R4	Q8c8r4 mus musculu
9	1672	84.5	380	2	Q8AWB4	Q8awb4 gallus gall
10	1586	80.1	351	2	Q7YRE6	Q7yre6 bos taurus
11	1490	75.3	380	2	Q640S0	Q640s0 xenopus tro
12	1490	75.3	396	2	Q6EV71	Q6ev71 xenopus tro
13	1474	74.5	395	2	Q7T028	Q7t028 brachydanio
14	1447	73.1	380	2	Q8AXS8	Q8axs8 xenopus lae
15	1427	72.1	384	2	Q70AG7	Q70ag7 fugu rubrip
16	1395	70.5	387	2	Q6EV65	Q6ev65 tetraodon n
17	848.5	42.9	381	2	Q6EV67	Q6ev67 ciona savig
18	838	42.3	402	2	Q659S1	Q659s1 drosophila
19	831	42.0	402	1	OFU1_DROME	Q9v6x7 drosophila
20	827	41.8	402	2	Q86SA7	Q86sa7 drosophila
21	789.5	39.9	399	2	Q7QHS7	Q7qhs7 anopheles g
22	786	39.7	353	2	Q659S0	Q659s0 bombyx mori
23	778	39.3	364	2	Q8MSR1	Q8msr1 drosophila
24	761	38.5	402	2	Q6EV68	Q6ev68 drosophila
25	660.5	33.4	381	1	OFU1_CAEEL	Q18014 caenorhabdi
26	644	32.5	434	2	Q659S2	Q659s2 caenorhabdi
27	578.5	29.2	385	2	Q659R9	Q659r9 onchocerca
28	295	14.9	61	1	OFU1_CRIGR	P83337 cricetulus
29	182.5	9.2	424	1	OFU2_CAEEL	Q8wr51 caenorhabdi
30	161.5	8.2	490	2	Q659L6	Q65916 drosophila
31	159.5	8.1	423	2	Q7K729	Q7k729 drosophila
32	159.5	8.1	490	2	Q9W589	Q9w589 drosophila
33	156.5	7.9	424	2	Q6EV55	Q6ev55 caenorhabdi
34	153.5	7.8	386	2	Q6EV53	Q6ev53 fugu rubrip
35	149.5	7.6	421	2	Q6EV54	Q6ev54 drosophila
36	144	7.3	429	2	Q6EV56	Q6ev56 pan troglod
37	142	7.2	429	2	Q7Z4N0	Q7z4n0 homo sapien
38	141	7.1	429	1	OFU2_MOUSE	Q8vhi3 mus musculu
39	137.5	6.9	434	2	Q7YZS7	Q7yzs7 ciona intes
40	136.5	6.9	380	2	Q7T1N6	Q7t1n6 gallus gall
41	134	6.8	403	2	Q6EV52	Q6ev52 ciona savig
42	130	6.6	429	2	Q7YRR5	Q7yrr5 bos taurus
43	127	6.4	372	2	Q7PRQ3	Q7prq3 anopheles g
44	125.5	6.3	362	2	Q659L5	Q6591s bombyx mori
45	125	6.3	445	2	Q9MA87	Q9ma87 arabidopsis

ALIGNMENTS

RESULT 1
OFU1_HUMAN
ID OFU1_HUMAN STANDARD; PRT; 388 AA.
AC Q9H488; Q14685; Q9BW76;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)
DE (Peptide O-fucosyltransferase) (O-FucT-1).
GN Name=POFUT1; Synonyms=KIAA0180;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND INDUCTION.
RC TISSUE=Heart;
RX MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;
RA Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,
RA Haltiwanger R.S.;
RT "Modification of epidermal growth factor-like repeats with O-fucose:
RT molecular cloning and expression of a novel GDP-fucose protein O-
RT fucosyltransferase.";
RL J. Biol. Chem. 276:40338-40345(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual
RT curation of 330 KIAA cDNA clones.";
RL DNA Res. 9:99-106(2002).
RN [3]
RP SEQUENCE OF 46-388 FROM N.A. (ISOFORM 1).
RC TISSUE=Bone marrow;
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 3:17-24(1996).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052; DOI=10.1038/414865a;
RA Deloukas P., Matthews L.H., Ashurst J.L., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,

RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=97175972; PubMed=9023546;
RA Wang Y., Lee G.F., Kelley R.F., Spellman M.W.;
RT "Identification of a GDP-L-fucose:polypeptide fucosyltransferase and
RT enzymatic addition of O-linked fucose to EGF domains.";
RL Glycobiology 6:837-842(1996).
CC -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC O-glycosidic linkage to a conserved serine or threonine residue in
CC EGF domains.
CC -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC -!- COFACTOR: Manganese.
CC -!- PATHWAY: Glycosylation.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9H488-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9H488-2; Sequence=VSP_001809;
CC Note=No experimental confirmation available;
CC -!- TISSUE SPECIFICITY: Highly expressed in heart, brain, placenta,
CC lung, liver, skeletal muscle, kidney and pancreas.
CC -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
CC ---

CC ---
DR EMBL; AF375884; AAL09576.1; -.
DR EMBL; D80002; BAA11497.2; -.
DR EMBL; AL121897; CAC16424.1; -.
DR EMBL; BC000582; AAH00582.1; -.
DR Genew; HGNC:14988; POFUT1.
DR H-InvDB; HIX0015724; -.
DR MIM; 607491; -.
DR GO; GO:0005737; C:cytoplasm; ISS.
DR GO; GO:0030173; C:integral to Golgi membrane; NAS.
DR GO; GO:0046922; F:peptide-O-fucosyltransferase activity; TAS.
DR GO; GO:0009790; P:embryonic development; NAS.
DR GO; GO:0007219; P:Notch signaling pathway; ISS.
DR GO; GO:0016266; P:O-glycan processing; TAS.
DR GO; GO:0045449; P:regulation of transcription; NAS.
DR GO; GO:0006445; P:regulation of translation; NAS.
KW Alternative splicing; Fucose metabolism; Glycoprotein;
KW Glycosyltransferase; Manganese; Signal; Transferase.
FT SIGNAL 1 26 Potential.
FT CHAIN 27 388 GDP-fucose protein O-fucosyltransferase
FT 1.
FT CARBOHYD 62 62 N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD 160 160 N-linked (GlcNAc. . .) (Potential).
FT VARSPLIC 182 388 FSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTG
FT EAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA
FT SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQ
FT SVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILG
FT QADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRD
FT EF -> RENHSCVTLLFPR (in isoform 2).
FT /FTId=VSP_001809.
SQ SEQUENCE 388 AA; 43955 MW; 3FACCCA434D02415 CRC64;

Query Match 100.0%; Score 1979; DB 1; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.1e-160;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 83

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         84 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 143

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 203

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 263

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 323

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 383

Qy        361 LRDEF 365
              |||||
Db        384 LRDEF 388
```

RESULT 2
OFU1_PANTR
ID OFU1_PANTR STANDARD; PRT; 388 AA.
AC Q6EV69;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1130

Qy        361 LeuArgAspGluPhe 365
                  ||||||||||||||||
Db       1131 CTGCGGGACGAGTTC 1145

RESULT 10
D80002
LOCUS       D80002                  5189 bp    mRNA    linear   PRI 10-JAN-2003
DEFINITION  Homo sapiens KIAA0180 mRNA, complete cds.
ACCESSION   D80002
VERSION     D80002.2  GI:20521837
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nagase,T., Seki,N., Ishikawa,K., Tanaka,A. and Nomura,N.
  TITLE     Prediction of the coding sequences of unidentified human genes. V.
            The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
            analysis of cDNA clones from human cell line KG-1
  JOURNAL   DNA Res. 3 (1), 17-24 (1996)
  MEDLINE   96281124
   PUBMED   8724849
REFERENCE   2
  AUTHORS   Chiang,P.W., Wang,S., Smithivas,P., Song,W.J., Ramamoorthy,S.,
            Hillman,J., Puett,S., Van Keuren,M.L., Crombez,E., Kumar,A.,
            Glover,T.W., Miller,D.E., Tsai,C.H., Blackburn,C.C., Chen,X.N.,
            Sun,Z., Cheng,J.F., Korenberg,J.R. and Kurnit,D.M.
  TITLE     Identification and analysis of the human and murine putative
            chromatin structure regulator SUPT6H and Supt6h
  JOURNAL   Genomics 34 (3), 328-333 (1996)
  MEDLINE   96374824
   PUBMED   8786132
REFERENCE   3  (bases 1 to 5189)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1995) Osamu Ohara, Kazusa DNA Research Institute;
            1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913)
COMMENT     On May 9, 2002 this sequence version replaced gi:1136419.
FEATURES             Location/Qualifiers
     source          1. .5189
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="ha02567s1"
                     /sex="male"
                     /cell_line="KG-1"
                     /cell_type="myeloblast"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
                     /note="This sequence was obtained by subcloning of the DNA
                     fragments derived from two cDNA clones (1 - 1066 was
                     derived from pf00933 and 1067 - 5189 was derived from
                     ha02567)."
     gene            1. .5189
                     /gene="KIAA0180"
     CDS             <1. .1213
                     /gene="KIAA0180"
                     /citation=[2]
                     /codon_start=2
                     /protein_id="BAA11497.2"
                     /db_xref="GI:20521838"
                     /translation="PPRLCAAAGSGSRADMGAAAWARPLSVSFLLLLLPLPGMPAGSW
                     DPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSY
                     QKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEG
                     NPFGPFWDQPHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE
                     EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHF
                     MASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPEL
                     QQLFKGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFG
                     MDRPPKLRDEF"
ORIGIN

Alignment Scores:
Pred. No.:                6.7e-174        Length:          5189
Score:                    1979.00         Matches:         365
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       9               Gaps:            0

US-09-774-954-2 (1-365) x D80002 (1-5189)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 175

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 235

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 295

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 355

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        356 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 415

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 475

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        476 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 535

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        536 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 595

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        596 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 655

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        656 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 715

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 775

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 835

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 895

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 955

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1015
```

```
Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1075

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1135

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1136 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1195

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1196 CTGCGGGACGAGTTC 1210
```

```
RESULT 11
AF375884
LOCUS       AF375884                5249 bp    mRNA    linear   PRI 23-OCT-2001
DEFINITION  Homo sapiens protein o-fucosyltransferase (POFUT1) mRNA, complete
            cds.
ACCESSION   AF375884
VERSION     AF375884.1  GI:15825113
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5249)
  AUTHORS   Wang,Y., Shao,L., Shi,S., Harris,R.J., Spellman,M.W., Stanley,P.
            and Haltiwanger,R.S.
  TITLE     Modification of epidermal growth factor-like repeats with O-fucose.
            Molecular cloning and expression of a novel GDP-fucose protein
            O-fucosyltransferase
  JOURNAL   J. Biol. Chem. 276 (43), 40338-40345 (2001)
  MEDLINE   21523965
   PUBMED   11524432
REFERENCE   2  (bases 1 to 5249)
  AUTHORS   Shi,S., Stanley,P., Wang,Y., Shao,L., Harris,R.J., Spellman,M.W.
            and Haltiwanger,R.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2001) Cell Biology, Albert Einstein College of
            Medicine, 1300 Morris Park Ave., New York, NY 10461, USA
FEATURES             Location/Qualifiers
     source          1. .5249
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="between PLAGL2 and KIF3B"
                     /tissue_type="heart"
     gene            1. .5249
                     /gene="POFUT1"
     CDS             50. .1216
                     /gene="POFUT1"
                     /function="catalyzes transfer of fucose from GDP-fucose to
                     serine or threonine on epidermal growth factor-like
                     repeats"
                     /note="O-FucT-1; glycosyltransferase"
                     /codon_start=1
                     /product="protein o-fucosyltransferase"
                     /protein_id="AAL09576.1"
                     /db_xref="GI:15825114"
                     /translation="MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRF
                     GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRV
                     ISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFN
                     KSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEM
                     VKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA
                     PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKP
                     EVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF"
ORIGIN
```

Alignment Scores:

```
Pred. No.:              6.8e-174        Length:         5249
Score:                  1979.00         Matches:        365
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-774-954-2 (1-365) x AF375884 (1-5249)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 598

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 658

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 718

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 778

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 838

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 898

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 958

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1018

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1078
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:21:23 ; Search time 210.606 Seconds
(without alignments)
887.482 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGYLLLYCPCMGR.....QGRPSSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1979	100.0	388	1	OFU1_HUMAN	Q9h488 homo sapien
2	1979	100.0	388	1	OFU1_PANTR	Q6ev69 pan troglod
3	1895	95.8	391	2	Q7YRE7	Q7yre7 bos taurus
4	1878	94.9	391	2	Q6EV66	Q6ev66 bos taurus
5	1854	93.7	395	2	Q6EV70	Q6ev70 rattus norv
6	1853	93.6	378	2	Q7YRZ3	Q7yrz3 sus scrofa
7	1832	92.6	393	1	OFU1_MOUSE	Q91zw2 mus musculu
8	1829	92.4	393	2	Q8C8R4	Q8c8r4 mus musculu
9	1672	84.5	380	2	Q8AWB4	Q8awb4 gallus gall
10	1586	80.1	351	2	Q7YRE6	Q7yre6 bos taurus
11	1490	75.3	380	2	Q640S0	Q640s0 xenopus tro
12	1490	75.3	396	2	Q6EV71	Q6ev71 xenopus tro
13	1474	74.5	395	2	Q7T028	Q7t028 brachydanio
14	1447	73.1	380	2	Q8AXS8	Q8axs8 xenopus lae
15	1427	72.1	384	2	Q70AG7	Q70ag7 fugu rubrip
16	1395	70.5	387	2	Q6EV65	Q6ev65 tetraodon n
17	848.5	42.9	381	2	Q6EV67	Q6ev67 ciona savig
18	838	42.3	402	2	Q659S1	Q659s1 drosophila
19	831	42.0	402	1	OFU1_DROME	Q9v6x7 drosophila
20	827	41.8	402	2	Q86SA7	Q86sa7 drosophila
21	789.5	39.9	399	2	Q7QHS7	Q7qhs7 anopheles g
22	786	39.7	353	2	Q659S0	Q659s0 bombyx mori
23	778	39.3	364	2	Q8MSR1	Q8msr1 drosophila
24	761	38.5	402	2	Q6EV68	Q6ev68 drosophila
25	660.5	33.4	381	1	OFU1_CAEEL	Q18014 caenorhabdi
26	644	32.5	434	2	Q659S2	Q659s2 caenorhabdi
27	578.5	29.2	385	2	Q659R9	Q659r9 onchocerca
28	295	14.9	61	1	OFU1_CRIGR	P83337 cricetulus
29	182.5	9.2	424	1	OFU2_CAEEL	Q8wr51 caenorhabdi
30	161.5	8.2	490	2	Q659L6	Q65916 drosophila
31	159.5	8.1	423	2	Q7K729	Q7k729 drosophila
32	159.5	8.1	490	2	Q9W589	Q9w589 drosophila
33	156.5	7.9	424	2	Q6EV55	Q6ev55 caenorhabdi
34	153.5	7.8	386	2	Q6EV53	Q6ev53 fugu rubrip
35	149.5	7.6	421	2	Q6EV54	Q6ev54 drosophila
36	144	7.3	429	2	Q6EV56	Q6ev56 pan troglod
37	142	7.2	429	2	Q7Z4N0	Q7z4n0 homo sapien
38	141	7.1	429	1	OFU2_MOUSE	Q8vhi3 mus musculu
39	137.5	6.9	434	2	Q7YZS7	Q7yzs7 ciona intes
40	136.5	6.9	380	2	Q7T1N6	Q7t1n6 gallus gall
41	134	6.8	403	2	Q6EV52	Q6ev52 ciona savig
42	130	6.6	429	2	Q7YRR5	Q7yrr5 bos taurus
43	127	6.4	372	2	Q7PRQ3	Q7prq3 anopheles g
44	125.5	6.3	362	2	Q659L5	Q65915 bombyx mori
45	125	6.3	445	2	Q9MA87	Q9ma87 arabidopsis

ALIGNMENTS

RESULT 1
OFU1_HUMAN
ID OFU1_HUMAN STANDARD; PRT; 388 AA.
AC Q9H488; Q14685; Q9BW76;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)
DE (Peptide O-fucosyltransferase) (O-FucT-1).
GN Name=POFUT1; Synonyms=KIAA0180;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND INDUCTION.
RC TISSUE=Heart;
RX MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;
RA Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,
RA Haltiwanger R.S.;
RT "Modification of epidermal growth factor-like repeats with O-fucose:
RT molecular cloning and expression of a novel GDP-fucose protein O-
RT fucosyltransferase.";
RL J. Biol. Chem. 276:40338-40345(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual
RT curation of 330 KIAA cDNA clones.";
RL DNA Res. 9:99-106(2002).
RN [3]
RP SEQUENCE OF 46-388 FROM N.A. (ISOFORM 1).
RC TISSUE=Bone marrow;
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 3:17-24(1996).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052; DOI=10.1038/414865a;
RA Deloukas P., Matthews L.H., Ashurst J.L., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,

```
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [6]
RP   CHARACTERIZATION.
RX   MEDLINE=97175972; PubMed=9023546;
RA   Wang Y., Lee G.F., Kelley R.F., Spellman M.W.;
RT   "Identification of a GDP-L-fucose:polypeptide fucosyltransferase and
RT   enzymatic addition of O-linked fucose to EGF domains.";
RL   Glycobiology 6:837-842(1996).
CC   -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC       O-glycosidic linkage to a conserved serine or threonine residue in
CC       EGF domains.
CC   -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC       beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC   -!- COFACTOR: Manganese.
CC   -!- PATHWAY: Glycosylation.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9H488-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9H488-2; Sequence=VSP_001809;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Highly expressed in heart, brain, placenta,
CC       lung, liver, skeletal muscle, kidney and pancreas.
CC   -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF375884; AAL09576.1; -.
DR   EMBL; D80002; BAA11497.2; -.
DR   EMBL; AL121897; CAC16424.1; -.
DR   EMBL; BC000582; AAH00582.1; -.
DR   Genew; HGNC:14988; POFUT1.
DR   H-InvDB; HIX0015724; -.
DR   MIM; 607491; -.
DR   GO; GO:0005737; C:cytoplasm; ISS.
DR   GO; GO:0030173; C:integral to Golgi membrane; NAS.
DR   GO; GO:0046922; F:peptide-O-fucosyltransferase activity; TAS.
DR   GO; GO:0009790; P:embryonic development; NAS.
DR   GO; GO:0007219; P:Notch signaling pathway; ISS.
DR   GO; GO:0016266; P:O-glycan processing; TAS.
DR   GO; GO:0045449; P:regulation of transcription; NAS.
DR   GO; GO:0006445; P:regulation of translation; NAS.
KW   Alternative splicing; Fucose metabolism; Glycoprotein;
KW   Glycosyltransferase; Manganese; Signal; Transferase.
FT   SIGNAL        1     26       Potential.
FT   CHAIN        27    388       GDP-fucose protein O-fucosyltransferase
FT                                 1.
FT   CARBOHYD     62     62       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    160    160       N-linked (GlcNAc. . .) (Potential).
FT   VARSPLIC    182    388       FSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTG
FT                                EAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA
FT                                SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQ
FT                                SVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILG
FT                                QADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRD
FT                                EF -> RENHSCVTLLFPR (in isoform 2).
FT                                /FTId=VSP_001809.
SQ   SEQUENCE   388 AA;  43955 MW;  3FACCCA434D02415 CRC64;
```

Query Match 100.0%; Score 1979; DB 1; Length 388;
Best Local Similarity 100.0%; Pred. No. 1.1e-160;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 83

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         84 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 143

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 203

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 263

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 323

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 383

Qy        361 LRDEF 365
              |||||
Db        384 LRDEF 388
```

```
RESULT 2
OFU1_PANTR
ID   OFU1_PANTR     STANDARD;      PRT;   388 AA.
AC   Q6EV69;
DT   25-OCT-2004 (Rel. 45, Created)
DT   25-OCT-2004 (Rel. 45, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)
```

```
DE   (Peptide O-fucosyltransferase) (O-FucT-1).
GN   Name=POFUT1; Synonyms=FUT12;
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=12966037; DOI=10.1093/glycob/cwg113;
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RA   Oriol R.;
RT   "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT   fucosyltransferases and alpha6-fucosyltransferases: phylogeny and
RT   identification of conserved peptide motifs.";
RL   Glycobiology 13:1C-5C(2003).
CC   -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC       O-glycosidic linkage to a conserved serine or threonine residue in
CC       EGF domains (By similarity).
CC   -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC       beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC   -!- COFACTOR: Manganese (By similarity).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ781500; CAH03712.1; -.
KW   Fucose metabolism; Glycoprotein; Glycosyltransferase; Manganese;
KW   Signal; Transferase.
FT   SIGNAL       1     26       Potential.
FT   CHAIN       27    388       GDP-fucose protein O-fucosyltransferase
FT                               1.
FT   CARBOHYD     62     62       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    160    160       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   388 AA;  43955 MW;  3FACCCA434D02415 CRC64;

  Query Match             100.0%;  Score 1979;  DB 1;  Length 388;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-160;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 83

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         84 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 143

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 203

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 263

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 323

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 383

Qy        361 LRDEF 365
              |||||
Db        384 LRDEF 388
```

```
RESULT 3
Q7YRE7
ID   Q7YRE7      PRELIMINARY;      PRT;   391 AA.
AC   Q7YRE7;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Protein O-fucosyltransferase 1a.
GN   Name=pofut1;
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Loriol C., Germot A., Dupuy F., Maftah A.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY344580; AAQ02332.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   391 AA;  44307 MW;  96762AB81A2027AD CRC64;

  Query Match             95.8%;  Score 1895;  DB 2;  Length 391;
  Best Local Similarity   93.4%;  Pred. No. 1.8e-153;
  Matches  341;  Conservative   17;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              :| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         27 LPVGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNV 86

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 146

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              |||||||||||||||||||||| |||||||::|| |||||:|||||||||||||||||||
Db        147 EGNPFGPFWDQFHVSFNKSELFAGISFSASYKDQWIQRFSPEEHPVLALPGAPAQFPVLE 206

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              |||||||||||||||:||||||||:|||||||||||||||||||||||||||:||||
Db        207 EHRPLQKYMVWSDEMVRTGEAQIHAHLIRPYVGIHLRIGSDWKNACAMLKDGTAGAHFMA 266

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||| | |||||||||||||:||:|||| :|:||||:|||||||:||:||||
Db        267 SPQCVGYSRHTTAPLTMTMCLPDLKEIRRALKLWVTALNAQSVYIATDSESYLPEIQQLF 326

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||:|||||||||||||||||||||||||||||| ||||||||||||:
Db        327 KGKVKVVSLKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLHGRPSSFFGMDRPPQ 386

Qy        361 LRDEF 365
              |||||
Db        387 LRDEF 391
```

```
RESULT 4
Q6EV66
ID   Q6EV66      PRELIMINARY;      PRT;   391 AA.
AC   Q6EV66;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Protein-O-fucosyltransferase 1.
GN   Name=fut12;
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RX   PubMed=12966037; DOI=10.1093/glycob/cwg113;
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RA   Oriol R.;
RT   "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT   fucosyltransferases and alpha6-fucosyltransferases: phylogeny and
RT   identification of conserved peptide motifs.";
RL   Glycobiology 13:1C-5C(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Martinez-Duncker I., Oriol R., Mollicone R.;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ781503; CAH03714.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   391 AA;  44277 MW;  4E047D6BB5CDF7FA CRC64;

  Query Match            94.9%;  Score 1878;  DB 2;  Length 391;
  Best Local Similarity  92.3%;  Pred. No. 5e-152;
  Matches  337;  Conservative   20;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              :| |||:||||||:|||||||||||||||||||| |:||||||||||||||||||||:
Db         27 LPVGSWNPAGYLLFCPCMGRFGNQADHFLGSLAFAPLMNRTLAVPPWIEYQHHKPPFTNV 86

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 146

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              |||||||||||||||||||||| ||||||||::|| |||||:||||||||||||||||||
Db        147 EGNPFGPFWDQFHVSFNKSELFAGISFSASYKDQWIQRFSPEEHPVLALPGAPAQFPVLE 206

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              |||||||||||||||:||||||||||:||||||||||||||||||||||||:||||
Db        207 EHRPLQKYMVWSDEMVRTGEAQIHAHLIRPYVGIHLRIGSDWKNACAMLKDGTAGAHFMA 266

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||| | ||||||||||||||:||:|||| :|:|||||:|||||||:||:||||
Db        267 SPQCVGYSRHTTAPLTMTMCLPDLKEIRRALKLWVTALNAQSVYIATDSESYLPEIQQLF 326

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              |||||||||||||:|||||||||||||||||||||||||| ||||||||||||:
Db        327 KGKVKVVSLKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLHGRPSSFFGMDRPPQ 386

Qy        361 LRDEF 365
              |||||
Db        387 LRDEF 391


RESULT 5
Q6EV70
ID   Q6EV70      PRELIMINARY;      PRT;   395 AA.
AC   Q6EV70;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Protein-O-fucosyltransferase 1.
GN   Name=fut12;
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar;
RX   PubMed=12966037; DOI=10.1093/glycob/cwg113;
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RA   Oriol R.;
RT   "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT   fucosyltransferases and alpha6-fucosyltransferases: phylogeny and
RT   identification of conserved peptide motifs.";
RL   Glycobiology 13:1C-5C(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar;
RA   Martinez-Duncker I., Oriol R., Mollicone R.;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ781499; CAH03711.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   395 AA;  44983 MW;  2A85521E22F38920 CRC64;

  Query Match            93.7%;  Score 1854;  DB 2;  Length 395;
  Best Local Similarity  94.2%;  Pred. No. 5.7e-150;
  Matches  341;  Conservative    8;  Mismatches   13;  Indels    0;  Gaps    0;

Qy          4 GSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVS 63
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         34 GSWDLAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVS 93

Qy         64 YQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGN 123
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db         94 YQKYFKLEPLQAYHRVISLEDFMEKLAPFHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGN 153

Qy        124 PFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHR 183
              |||||||||||||||||||||||||||:||| ||| ||||||||||||||||||||||||
Db        154 PFGPFWDQFHVSFNKSELFTGISFSASYKEQWIQRFPPKEHPVLALPGAPAQFPVLEEHR 213

Qy        184 PLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQ 243
              |||||||||||:|||||| ||||||||||||||||||||||||||||||||||||||||
Db        214 ELQKYMVWSDEMVRTGEAQISTHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQ 273

Qy        244 CVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGK 303
              ||||||||| ||| |||||||||||||||||:|||:|||:|||||||| |:||||| |
Db        274 CVGYSRSTATPLTTTMCLPDLKEIQRAVKLWVRALDARSVYIATDSESYVSEIQQLFKEK 333

Qy        304 VKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRD 363
              |||||||||||:|||||||||||||||||||||||| |||||||||||| :|||
Db        334 VKVVSLKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLHGRPSSFFGMDRPSQLRD 393

Qy        364 EF 365
              ||
Db        394 EF 395


RESULT 6
Q7YRZ3
ID   Q7YRZ3      PRELIMINARY;      PRT;   378 AA.
AC   Q7YRZ3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Protein-O-fucosyltransferase (Fragment).
GN   Name=POFUT1;
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Martinez-Dunker I., Mollicone R., Candelier J.J., Breton C., Oriol R.;
RT   "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT   fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and
RT   identification of conserved peptide motifs.";
RL   Glycobiology 13:1c-5c(2003).
DR   EMBL; AJ567917; CAD99201.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     378    378
SQ   SEQUENCE   378 AA;  42842 MW;  1870DEB0C50713F1 CRC64;

  Query Match            93.6%;  Score 1853;  DB 2;  Length 378;
  Best Local Similarity  93.6%;  Pred. No. 6.6e-150;
  Matches  336;  Conservative   14;  Mismatches    9;  Indels    0;  Gaps    0;
```

```
Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              :|| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 LPAVSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 79

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              |||||||||||||| |||||||||||||||||||||:|||||||||||||||||||||||
Db         80 HVSYQKYFKLEPLQVYHRVISLEDFMEKLAPTHWPPERRVAYCFEVAAQRSPDKKTCPMK 139

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||| |||||||:|| ||||||||||||||||||||||||||||
Db        140 EGNPFGPFWDQFHVSFNKSELFAGISFSASYKDQWIQRFSPKEHPVLALPGAPAQFPVLE 199

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              |||||||||||||||:|||||| |||:|||||||||||||||||||||||:||||||||||
Db        200 EHRPLQKYMVWSDEMVRTGEAQIRAHLIRPYVGIHLRIGSDWKNACAMLKDGTAGAHFMA 259

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||| |||||||||||||||:||:|||| :| |||||:|||||||:||:||||||
Db        260 SPQCVGYSRHTAAPLTMTMCLPDLKEIRRALKLWVTALSAQSVYIATDSESYLPEIQQLF 319

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPP 359
              |||||||||||||||:|||||||||||||||||||||||||||||| ||||||||||||
Db        320 KGKVKVVSLKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLQGRLSSFFGMDRPP 378
```

RESULT 7
OFU1_MOUSE
ID OFU1_MOUSE STANDARD; PRT; 393 AA.
AC Q91ZW2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)
DE (Peptide O-fucosyltransferase) (O-FucT-1).
GN Name=Pofut1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SvJ; TISSUE=Liver;
RX MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;
RA Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,
RA Haltiwanger R.S.;
RT "Modification of epidermal growth factor-like repeats with O-fucose:
RT molecular cloning and expression of a novel GDP-fucose protein O-
RT fucosyltransferase.";
RL J. Biol. Chem. 276:40338-40345(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N-3; TISSUE=Mammary gland;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC   O-glycosidic linkage to a conserved serine or threonine residue in
CC   EGF domains (By similarity).
CC -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC   beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC -!- COFACTOR: Manganese (By similarity).
CC -!- PATHWAY: Glycosylation.
CC -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
CC ---

CC ---
DR EMBL; AF375885; AAL09577.1; -.
DR EMBL; BC046295; AAH46295.1; -.
DR MGD; MGI:2153207; Pofut1.
DR GO; GO:0030173; C:integral to Golgi membrane; ISS.
DR GO; GO:0008417; F:fucosyltransferase activity; ISS.
DR GO; GO:0001525; P:angiogenesis; IMP.
DR GO; GO:0007507; P:heart development; IMP.
DR GO; GO:0007399; P:neurogenesis; IMP.
DR GO; GO:0007219; P:Notch signaling pathway; IMP.
DR GO; GO:0006493; P:O-linked glycosylation; ISS.
DR GO; GO:0001756; P:somitogenesis; IMP.
KW Fucose metabolism; Glycoprotein; Glycosyltransferase; Manganese;
KW Signal; Transferase.
FT SIGNAL   1   30   Potential.
FT CHAIN   31   393   GDP-fucose protein O-fucosyltransferase
FT   1.
FT CARBOHYD   67   67   N-linked (GlcNAc. . .) (Potential).
FT CARBOHYD   165   165   N-linked (GlcNAc. . .) (Potential).
SQ SEQUENCE   393 AA; 44688 MW; D982104E95E5CF3B CRC64;

  Query Match   92.6%; Score 1832; DB 1; Length 393;
  Best Local Similarity   92.6%; Pred. No. 4.3e-148;
  Matches 336; Conservative   13; Mismatches   14; Indels   0; Gaps   0;

```
Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHV 62
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 AGSWDLAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHV 90

Qy         63 SYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEG 122
              |||||||||||||||:||||||| |||:|||||||||||||||||||||||||||||||||
Db         91 SYQKYFKLEPLQAYHRVVSLEDFMENLAPSHWPPEKRVAYCFEVAAQRSPDKKTCPMKEG 150

Qy        123 NPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEH 182
              ||||||||||||||||||||||||||:|||:|||  ||||||||||||||||||||||||
Db        151 NPFGPFWDQFHVSFNKSELFTGISFSASYKEQWTQRFPAKEHPVLALPGAPAQFPVLEEH 210

Qy        183 RPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASP 242
              | ||||||||||||:|||| | |||||||||||||||||||||||||||||||||||||||
Db        211 RELQKYMVWSDEMVRTGEALISAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASP 270

Qy        243 QCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKG 302
              |||||||||| |||||||||||||||||| ||||:|:|:|||:|||||||| |:|||||
Db        271 QCVGYSRSTATPLTMTMCLPDLKEIQRAVTLWVRALNARSVYIATDSESYVSEIQQLFKD 330

Qy        303 KVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLR 362
              ||:|||||||||:|||||||||||||||||||||||||||| || ||||||||| :||
Db        331 KVRVVSLKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLHGRQSSFFGMDRPSQLR 390

Qy        363 DEF 365
              |||
Db        391 DEF 393
```

RESULT 8
Q8C8R4

```
ID   Q8C8R4      PRELIMINARY;      PRT;   393 AA.
AC   Q8C8R4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Mus musculus adult retina cDNA, RIKEN full-length enriched library,
DE   clone:A930028F21 product:protein O-fucosyltransferase 1, full insert
DE   sequence.
GN   Name=Pofut1;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK044629; BAC32009.1; -.
DR   MGD; MGI:2153207; Pofut1.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0001525; P:angiogenesis; IMP.
DR   GO; GO:0006004; P:fucose metabolism; TAS.
DR   GO; GO:0007507; P:heart development; IMP.
DR   GO; GO:0007399; P:neurogenesis; IMP.
DR   GO; GO:0007219; P:Notch signaling pathway; IMP.
DR   GO; GO:0001756; P:somitogenesis; IMP.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   393 AA;  44704 MW;  1CA94CB30B85D6DC CRC64;

  Query Match           92.4%;  Score 1829;  DB 2;  Length 393;
  Best Local Similarity   92.3%;  Pred. No. 7.8e-148;
  Matches  335;  Conservative   14;  Mismatches   14;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHV 62
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 AGSWDLAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHV 90

Qy         63 SYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEG 122
              ||||||||||||||||||:||||||| |||:|||||||||||||||||||||||||||||
Db         91 SYQKYFKLEPLQAYHRVVSLEDFMENLAPSHWPPEKRVAYCFEVAAQRSPDKKTCPMKEG 150

Qy        123 NPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEH 182
              ||||||||||||||||||||||||||||:|||:|||  ||||||||||||||||||||||
Db        151 NPFGPFWDQFHVSFNKSELFTGISFSASYKEQWTQRFPAKEHPVLALPGAPAQFPVLEEH 210

Qy        183 RPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASP 242
              | |||||||||||:|||| | :||||||||||||||||||||||||||||||||||||||
Db        211 RELQKYMVWSDEMVRTGEALISSHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASP 270

Qy        243 QCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKG 302
              |||||||||| |||||||||||||||| ||||:|:|:|||:|||||||| |:|||||
Db        271 QCVGYSRSTATPLTMTMCLPDLKEIQRAVTLWVRALNARSVYIATDSESYVSEIQQLFKD 330

Qy        303 KVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLR 362
              ||:||||||||:|||||||||||||||||||||||||||| || ||||||||| :||
Db        331 KVRVVSLKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLHGRQSSFFGMDRPSQLR 390

Qy        363 DEF 365
              |||
Db        391 DEF 393


RESULT 9
Q8AWB4
ID   Q8AWB4      PRELIMINARY;      PRT;   380 AA.
AC   Q8AWB4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein POFUT1.
GN   Name=POFUT1;
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Oriol R.;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ535754; CAD59738.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   380 AA;  43316 MW;  83B53E342511811B CRC64;

  Query Match           84.5%;  Score 1672;  DB 2;  Length 380;
  Best Local Similarity   81.5%;  Pred. No. 2e-134;
  Matches  296;  Conservative   33;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHV 62
              ||||| |||||||||||||||:||||:||||: |||||||||||||:||:|:|||||
```

```
Db         18 AGSWDTAGYLLYCPCMGRFGNQAEHFLGALAFARALNRTLAVPPWIEYRHHRPPYTNLHV 77

Qy         63 SYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEG 122
              |::|||||||| ||||:||| |||:||| |||| :|||||||| ||||| || ||||||:|
Db         78 PYEEYFKLEPLQQYHRVLSLEQFMEQLAPQHWPPGRRVAYCFEAAAQRSADKSTCPMKDG 137

Qy        123 NPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEH 182
              |||||||||| | |:||||| |||||::||: | ||| | ||||||||||||||||||||
Db        138 NPFGPFWDQFKVDFDKSELFKGISFSSAYRDHWIQRFPPSEHPVLALPGAPAQFPVLEEH 197

Qy        183 RPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASP 242
              ||| |||||||||| ||| ||: |||||||||||||||||| ||||||||||||||||
Db        198 RPLHKYMVWSDEMVKKGEAYIHSLLVRPYVGIHLRIGSDWKNACNMLKDGTAGSHFMASP 257

Qy        243 QCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKG 302
              ||||| |||| |||| ||||||||||:||:|:||:   |:|:|:|||||  |  |:|| |:|
Db        258 QCVGYDRSTAVPLTMDMCLPDLKEIKRALKVWVKKTVAKSIYIATDSEPYTKEIQQFFQG 317

Qy        303 KVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLR 362
              |:|||||:|||||:|||||||:||||||||||||||||||: |:|||||| ||:||
Db        318 KIKVVSLQPEVAQIDLYILGQSDHFIGNCVSSFTAFVKRERDVHGKPSSFFGMDHPPRLR 377

Qy        363 DEF 365
              |||
Db        378 DEF 380


RESULT 10
Q7YRE6
ID   Q7YRE6      PRELIMINARY;      PRT;   351 AA.
AC   Q7YRE6;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Protein O-fucosyltransferase 1b.
GN   Name=pofut1;
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Loriol C., Germot A., Dupuy F., Maftah A.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY344581; AAQ02333.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   351 AA;  39576 MW;  BF830F61A7296F42 CRC64;

  Query Match             80.1%;  Score 1586;  DB 2;  Length 351;
  Best Local Similarity   88.6%;  Pred. No. 4e-127;
  Matches  288;  Conservative   20;  Mismatches    9;  Indels    8;  Gaps    3;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              :| |||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         27 LPVGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNV 86

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         87 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 146

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              |||||||||||||||||||| |||||||::|| |||||:|||||||||||||||||||
Db        147 EGNPFGPFWDQFHVSFNKSELFAGISFSASYKDQWIQRFSPEEHPVLALPGAPAQFPVLE 206

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              |||||||||||||||:||||||||:|||||||||||||||||||||||||||||||||||
Db        207 EHRPLQKYMVWSDEMVRTGEAQIHAHLIRPYVGIHLRIGSDWKNACAMLKDGTAGAHFMA 266

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||| | |||||||||||||:||:|||| :|:|||||:|||||||:||:||||||
Db        267 SPQCVGYSRHTTAPLTMTMCLPDLKEIRRALKLWVTALNAQSVYIATDSESYLPEIQQLF 326

Qy        301 KGKVKVVSLK---PEVAQVDLYILG 322
              |||    ||    |: ::: | :||
Db        327 KGK----SLSFAVPKASEMSL-LLG 346


RESULT 11
Q640S0
ID   Q640S0      PRELIMINARY;      PRT;   380 AA.
AC   Q640S0;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein.
OS   Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8364;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Klein S., Gerhard D.S.;
RL   Submitted (SEP-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC082519; AAH82519.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   380 AA;  43570 MW;  8BB938CE776D5238 CRC64;

  Query Match             75.3%;  Score 1490;  DB 2;  Length 380;
  Best Local Similarity   72.7%;  Pred. No. 7.2e-119;
  Matches  261;  Conservative   48;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          6 WDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQ 65
              || :||:|||||||||||||||||||::|||| ||||| | ||:||:||:|| |:
Db         21 WDSSGYILYCPCMGRFGNQADHFLGSLAFAKMVNRTLVVPPWIVYNHHRPPYTNVHVPYE 80

Qy         66 KYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPF 125
              ::|:|||| :| |||:| |||:||   ||| :|||||  |||||||||:||||:||||
Db         81 EFFQLEPLRQFHHVISMEQFMEQLASVLWPPSRRVAYCFTAAAQRSPDKKSCPMKDGNPF 140

Qy        126 GPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPL 185
              ||||| |||:| :|||| ||:||| |:: |: || | :|||:||||||||||||||| |
Db        141 GPFWDHFHVNFIQSELFDGITFSAYYKDVWNSRFPPSQHPVIALPGAPAQFPVLEEHRSL 200

Qy        186 QKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCV 245
              |:|:|||:::|: || || : |||||||:|||||||||| ||||||||||||||||||||
Db        201 QRYVVWSEKIVRAGEEQIQSLLVRPYVGVHLRIGSDWKNACKMLKDGTAGSHFMASPQCV 260

Qy        246 GYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVK 305
```

```
Db        261 GYDKYRAAPLTMEMCLPDLKEMRRALTLWVERSKARSVYIATDSTSHTEEIQAFLGDKVR 320

Qy        306 VVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDE 364

Db        321 VVCLQPEVAQMDLYILGQADHFIGNCVSSFTAFVKRERDVHGRPSSFFGMDSPGAVQDE 379


RESULT 12
Q6EV71
ID   Q6EV71      PRELIMINARY;      PRT;   396 AA.
AC   Q6EV71;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Protein-O-fucosyltransferase.
GN   Name=fut12;
OS   Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8364;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Martinez-Duncker I., Oriol R., Mollicone R.;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ781498; CAH03710.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR   InterPro; IPR000886; ER_target_S.
DR   PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   396 AA;  45056 MW;  F90504E19507073F CRC64;

  Query Match             75.3%;  Score 1490;  DB 2;  Length 396;
  Best Local Similarity   72.7%;  Pred. No. 7.6e-119;
  Matches  261;  Conservative   48;  Mismatches   50;  Indels    0;  Gaps    0;

Qy          6 WDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQ 65

Db         37 WDSSGYILYCPCMGRFGNQADHFLGSLAFAKMVNRTLVVPPWIVYNHHRPPYTNVHVPYE 96

Qy         66 KYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPF 125

Db         97 EFFQLEPLRQFHHVISMEQFMEQLASVLWPPSRRVAYCFTAAAQRSPDKKSCPMKDGNPF 156

Qy        126 GPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPL 185

Db        157 GPFWDHFHVNFIQSELFDGITFSAYYKDVWNSRFPPSQHPVIALPGAPAQFPVLEEHRSL 216

Qy        186 QKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCV 245

Db        217 QRYVVWSEKIVRAGEEQIQSLLVRPYVGVHLRIGSDWKNACKMLKDGTAGSHFMASPQCV 276

Qy        246 GYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVK 305

Db        277 GYDKYRAAPLTMEMCLPDLKEMRRALTLWVERSKARSVYIATDSTSHTEEIQAFLGDKVR 336

Qy        306 VVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDE 364

Db        337 VVCLQPEVAQMDLYILGQADHFIGNCVSSFTAFVKRERDVHGRPSSFFGMDSPGAVQDE 395


RESULT 13
Q7T028
ID   Q7T028      PRELIMINARY;      PRT;   395 AA.
AC   Q7T028;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   GDP-fucose protein O-fucosyltransferase 1 (EC 2.4.1.221).
GN   Name=pofut1; Synonyms=fut12;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RA   Oriol R.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ579536; CAE18459.1; -.
DR   ZFIN; ZDB-GENE-040303-2; pofut1.
DR   GO; GO:0046922; F:peptide-O-fucosyltransferase activity; IEA.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   395 AA;  45165 MW;  15BBE06172542E8A CRC64;

  Query Match             74.5%;  Score 1474;  DB 2;  Length 395;
  Best Local Similarity   73.9%;  Pred. No. 1.8e-117;
  Matches  263;  Conservative   40;  Mismatches   53;  Indels    0;  Gaps    0;

Qy          5 SWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSY 64

Db         33 TWDENGYILYCPCMGRFGNQVDHFLGSLAFAKMLNRTLAVPPWIVYRHHSPPFTNVHVPY 92

Qy         65 QKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNP 124

Db         93 SEYFQLEPLKQYHRVVSLEDFMEHLAPKHWPHGQRIAYCFESAAHRSQDKKSCPMKDGNP 152

Qy        125 FGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRP 184

Db        153 FGPFWDHIGVDFDRSVLFGGLSFSSYYQPHWIKRFPPKEHPVLALPGAPAQFPVLEEHIG 212

Qy        185 LQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQC 244

Db        213 LQQFVVWSDKIVQEGEGHIRNLLNRPYVGIHLRIGSDWQNACKLLKTGDTGPHFMASPQC 272

Qy        245 VGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKV 304

Db        273 VGYDRQTALPLTMNMCLPDLTEIRRAVKLWVKNTGAQSVYIATDSESHTTEIQKLFKGKV 332

Qy        305 KVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360

Db        333 KVVSLQPDTAQVDLYILGQADHFIGNCVSSFSAFVKRERNIHGRPSSFFGLDYPGK 388


RESULT 14
Q8AXS8
ID   Q8AXS8      PRELIMINARY;      PRT;   380 AA.
AC   Q8AXS8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein pofut1.
GN   Name=pofut1;
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Oriol R.;
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ514425; CAD55833.1; -.
DR   InterPro; IPR000886; ER_target_S.
DR   PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   380 AA;  43505 MW;  C6B960E76D875293 CRC64;

  Query Match             73.1%;  Score 1447;  DB 2;  Length 380;
  Best Local Similarity   70.6%;  Pred. No. 3.4e-115;
  Matches  255;  Conservative   50;  Mismatches   56;  Indels    0;  Gaps    0;

Qy          4 GSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVS 63
```

```
Db         19 GLYNSNGYIVYCPCMGRFGNQADHFLGSLAFAKMVNRTLVVPPWIVYNHHRPPYTNVHVP 78

Qy         64 YQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGN 123

Db         79 YEEFFQLEPLRQFHQVISMEEFMEQLAPVLWPPTKRLAYCFTAAAQRSPDKKSCPMKDGN 138

Qy        124 PFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHR 183

Db        139 PFGPFWDHFHVDFTQSELFDGITFSAYYKDVWISRFPPSQHPVIALPGAPAQFPVLEEHR 198

Qy        184 PLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQ 243

Db        199 SLQRYIVWSEKIVREGEEQIRSLLVRPYVGIHLRIGSDWKDACKMLKDGTAGSHVMASAQ 258

Qy        244 CVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGK 303

Db        259 CVCYDRYRAGPLTMDMCLPDLKEMRRALTLWVERSKARSVYIATDSPSYTAEIQAFLGNE 318

Qy        304 VKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRD 363

Db        319 VRVVCLQPEVAQIDLYILAQADHFIGNCVSSFTAFVKRDRDVHGRPSSFFGMDSPGVVQD 378

Qy        364 E 364

Db        379 E 379


RESULT 15
Q70AG7
ID   Q70AG7      PRELIMINARY;      PRT;   384 AA.
AC   Q70AG7;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Protein-O-fucosyltransferase.
GN   Name=fut12;
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetradontoidea; Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Oriol R.;
RL   Submitted (NOV-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ606070; CAE54305.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR   InterPro; IPR000886; ER_target_S.
DR   PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   384 AA;  43650 MW;  7CAA6ADC90F40032 CRC64;

  Query Match             72.1%;  Score 1427;  DB 2;  Length 384;
  Best Local Similarity   71.4%;  Pred. No. 1.8e-113;
  Matches  255;  Conservative   43;  Mismatches   59;  Indels    0;  Gaps    0;

Qy          6 WDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQ 65

Db         23 WDQNGYVLYCPCMGRFGNQADHFLGSLAFAKMLNRTLAVPPWIVYRHHVPPYTNVHIPYR 82

Qy         66 KYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPF 125

Db         83 DFFQLEALSAYHRVVSLEDFMETLAPRYWPAGQRRAYCFETAAQRTADKKSCPMKDGNPF 142

Qy        126 GPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPL 185

Db        143 GPFWDYVNVDFDESVLFGGIYFSAYYQPQWMKKFPPSQHPVLVLPGAPAQFPVSEEHVGL 202

Qy        186 QKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCV 245

Db        203 QRYMVWSEKMVEEGDGHIEKLLVRPYVGIHLRIGSDWQTACGMLDSGDMGPHFMASPQCV 262

Qy        246 GYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVK 305

Db        263 GYERQTALPLTATMCLPDLGEILRAVKVWVKKTSARSVYIATDSESHSGDIEQLFNGKVK 322

Qy        306 VVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLR 362

Db        323 VVSLRPELAQMDLYILGKADHFIGNCVSSFSAFVKRQRDVQGLPSSFFGMDTPGNLR 379


Search completed: October 25, 2005, 15:35:23
Job time : 213.606 secs
```

Organism DNA

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 05:08:04 ; Search time 8270.58 Seconds
(without alignments)
2138.440 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....QGRPSSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105432_6389/app_query.fasta_1.917
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954_@CGN_1_1_4968_@runat_25102005_105432_6389 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1979	100.0	1100	6	AR163457	AR163457 Sequence
2	1979	100.0	1100	6	BD103196	BD103196 O-fucosyl
3	1979	100.0	1167	6	CQ727777	CQ727777 Sequence
4	1979	100.0	1167	9	AJ781500	AJ781500 Pan trogl
5	1979	100.0	1300	6	AR163452	AR163452 Sequence
6	1979	100.0	1300	6	BD103189	BD103189 O-fucosyl
7	1979	100.0	1514	6	AR163451	AR163451 Sequence
8	1979	100.0	1514	6	BD103188	BD103188 O-fucosyl
9	1979	100.0	4560	9	HSM803330	AL832023 Homo sapi
10	1979	100.0	5189	9	D80002	D80002 Homo sapien
11	1979	100.0	5249	9	AF375884	AF375884 Homo sapi
12	1979	100.0	5266	6	AX780153	AX780153 Sequence
13	1979	100.0	11284	6	BD103190	BD103190 O-fucosyl
14	1895	95.8	1176	4	AY344580	AY344580 Bos tauru
15	1878	94.9	1176	4	AJ781503	AJ781503 Bos tauru
16	1854	93.7	1188	10	AJ781499	AJ781499 Rattus no
17	1853	93.6	1134	4	SSC567917	AJ567917 Sus scrof
18	1844	93.2	5009	6	AR163453	AR163453 Sequence
19	1844	93.2	5009	6	BD103191	BD103191 O-fucosyl
20	1832	92.6	1302	10	AF375885	AF375885 Mus muscu
21	1832	92.6	1529	10	BC046295	BC046295 Mus muscu
22	1672	84.5	1732	5	GGA535754	AJ535754 Gallus ga
23	1672	84.5	2546	5	AJ720352	AJ720352 Gallus ga
24	1586	80.1	1056	4	AY344581	AY344581 Bos tauru
25	1490	75.3	1502	5	BC082519	BC082519 Xenopus t
26	1490	75.3	1538	5	AJ781498	AJ781498 Silurana
27	1474	74.5	1549	5	DRE579536	AJ579536 Danio rer
28	1447	73.1	1505	5	XLA514425	AJ514425 Xenopus l
29	1427	72.1	1155	5	AJ606070	AJ606070 Fugu rubr
30	1395	70.5	1164	5	AJ781504	AJ781504 Tetraodon
31	1186	59.9	920	5	AJ719585	AJ719585 Gallus ga
32	889	44.9	1524	9	BC000582	BC000582 Homo sapi
33	870.5	44.0	1841	3	AK112708	AK112708 Ciona int
34	848.5	42.9	1146	3	AJ781502	AJ781502 Ciona sav
35	838	42.3	1209	3	AJ831490	AJ831490 Drosophil
36	837	42.3	1295	3	AY118651	AY118651 Drosophil
37	837	42.3	1320	6	AR509393	AR509393 Sequence
38	831	42.0	1209	6	CQ590403	CQ590403 Sequence
39	827	41.8	1579	3	AB093572	AB093572 Drosophil
c 40	796	40.2	3264	6	CQ590402	CQ590402 Sequence
c 41	796	40.2	41552	2	AC017347	AC017347 Drosophil
42	796	40.2	165177	3	AC007588	AC007588 Drosophil
c 43	796	40.2	186333	3	AC007852	AC007852 Drosophil
44	796	40.2	229684	3	AE003815	AE003815 Drosophil
45	786	39.7	1062	3	AJ831491	AJ831491 Bombyx mo

ALIGNMENTS

RESULT 1
AR163457
LOCUS AR163457 1100 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 16 from patent US 6270987.
ACCESSION AR163457
VERSION AR163457.1 GI:16234062
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1100)
AUTHORS Wang,Y. and Spellman,M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: US 6270987-A 16 07-AUG-2001;
FEATURES Location/Qualifiers
source 1. .1100
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Alignment Scores:

Pred. No.:	8.91e-175	Length:	1100
Score:	1979.00	Matches:	365
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-774-954-2 (1-365) x AR163457 (1-1100)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1081 CTGCGGGACGAGTTC 1095
```

RESULT 2
BD103196

```
LOCUS       BD103196                1100 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103196
VERSION     BD103196.1  GI:22648770
KEYWORDS    JP 2001527389-A/9.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 9 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/9
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .1100
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN
```

Alignment Scores:

Pred. No.:	8.91e-175	Length:	1100
Score:	1979.00	Matches:	365
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-774-954-2 (1-365) x BD103196 (1-1100)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
```

```
Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1081 CTGCGGGACGAGTTC 1095
```

```
RESULT 3
CQ727777
LOCUS       CQ727777               1167 bp    DNA     linear   PAT 03-FEB-2004
DEFINITION  Sequence 13711 from Patent WO02068579.
ACCESSION   CQ727777
VERSION     CQ727777.1  GI:42294771
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. .
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 13711 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .1167 .
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

Alignment Scores:

Pred. No.:	9.63e-175	Length:	1167
Score:	1979.00	Matches:	365
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-774-954-2 (1-365) x CQ727777 (1-1167)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 129

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 189

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 249

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 309

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 369

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 429

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 489

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 549

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 609

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        610 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 669

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        670 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 729

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        730 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 789

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        790 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 849

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        850 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 909

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        910 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 969

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        970 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1029

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1030 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1089

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1090 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1149

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1150 CTGCGGGACGAGTTC 1164

RESULT 4
AJ781500
LOCUS       AJ781500               1167 bp    mRNA    linear   PRI 07-JUL-2004
DEFINITION  Pan troglodytes mRNA for protein-O-fucosyltransferase 1 (fut12
            gene).
ACCESSION   AJ781500
VERSION     AJ781500.1  GI:50057069
KEYWORDS    fut12 gene; pofut1 gene; protein-O-fucosyltransferase 1.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Martinez-Duncker,I., Mollicone,R., Candelier,J.J., Breton,C. and
            Oriol,R.
  TITLE     A new superfamily of protein-O-fucosyltransferases,
            alpha2-fucosyltransferases, and alpha6-fucosyltransferases:
            phylogeny and identification of conserved peptide motifs
  JOURNAL   Glycobiology 13 (12), 1C-5C (2003)
   PUBMED   12966037
REFERENCE   2
  AUTHORS   Martinez-Duncker,I., Oriol,R. and Mollicone,R.
  TITLE     Phylogeny of fucosyltransferases
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1167)
  AUTHORS   Oriol,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Oriol R., U504, Inserm, 16 Av. Paul
            Vaillant-Couturier, 94807, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1167
                     /organism="Pan troglodytes"
                     /mol_type="mRNA"
                     /db_xref="taxon:9598"
     gene            1. .1167
                     /gene="fut12"
     CDS             1. .1167
                     /gene="fut12"
                     /function="adds fucose to Ser or Thr"
                     /note="pofut1"
                     /codon_start=1
                     /product="protein-O-fucosyltransferase 1"
                     /protein_id="CAH03712.1"
                     /db_xref="GI:50057070"
                     /translation="MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRF
                     GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRV
                     ISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFN
                     KSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEM
                     VKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA
                     PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKP
                     EVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF"
ORIGIN

Alignment Scores:
Pred. No.:            9.63e-175      Length:        1167

Score:                1979.00        Matches:       365
Percent Similarity:   100.00%        Conservative:  0
Best Local Similarity: 100.00%       Mismatches:    0
Query Match:          100.00%        Indels:        0
DB:                   9              Gaps:          0

US-09-774-954-2 (1-365) x AJ781500 (1-1167)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 129

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 189

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 249

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 309

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 369

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 429

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 489

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 549

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCGGCCCAGTTCCCCGTCCTGGAG 609

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        610 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 669

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        670 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 729

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        730 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACCGCAGGCTCGCACTTCATGGCC 789

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        790 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 849

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        850 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 909

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        910 CAGTCGGTCTATGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 969

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        970 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTTGACCTGTACATC 1029

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1030 CTCGGCCAAGCTGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1089

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1090 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1149

Qy        361 LeuArgAspGluPhe 365
                |||||||||||||||
Db       1150 CTGCGGGACGAGTTC 1164

RESULT 5
AR163452
LOCUS       AR163452               1300 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 6 from patent US 6270987.
ACCESSION   AR163452
VERSION     AR163452.1  GI:16234057
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1300)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: US 6270987-A 6 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1300
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:                1.11e-174      Length:          1300
Score:                    1979.00        Matches:         365
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       6              Gaps:            0

US-09-774-954-2 (1-365) x AR163452 (1-1300)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy        361 LeuArgAspGluPhe 365
                |||||||||||||||
Db       1216 CTGCGGGACGAGTTC 1230

RESULT 6
BD103189
LOCUS       BD103189               1300 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103189
VERSION     BD103189.1  GI:22648763
KEYWORDS    JP 2001527389-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1300)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 2 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/2
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .1300
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
```

```
     source          1. .1300
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

Alignment Scores:
Pred. No.:              1.11e-174       Length:         1300
Score:                  1979.00         Matches:        365
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-774-954-2 (1-365) x BD103189 (1-1300)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1216 CTGCGGGACGAGTTC 1230

RESULT 7
AR163451
LOCUS       AR163451                1514 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 2 from patent US 6270987.
ACCESSION   AR163451
VERSION     AR163451.1  GI:16234056
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: US 6270987-A 2 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1514
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              1.35e-174       Length:         1514
Score:                  1979.00         Matches:        365
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-774-954-2 (1-365) x AR163451 (1-1514)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
```

```
Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy        361 LeuArgAspGluPhe 365
Db       1081 CTGCGGGACGAGTTC 1095

RESULT 8
BD103188
LOCUS       BD103188                1514 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103188
VERSION     BD103188.1  GI:22648762
KEYWORDS    JP 2001527389-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 1 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/1
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .1514
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1514
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

Alignment Scores:
Pred. No.:                1.35e-174      Length:          1514
Score:                    1979.00        Matches:         365
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       6              Gaps:            0

US-09-774-954-2 (1-365) x BD103188 (1-1514)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy        361 LeuArgAspGluPhe 365
                |||||||||||||||
Db       1081 CTGCGGGACGAGTTC 1095

RESULT 9
HSM803330
LOCUS       HSM803330               4560 bp    mRNA    linear   PRI 13-MAY-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp451J114 (from clone DKFZp451J114).
ACCESSION   AL832023
VERSION     AL832023.1  GI:21732563
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4560)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp451J114) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1. .4560
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp451J114"
                     /db_xref="taxon:9606"
                     /clone="DKFZp451J114"
                     /tissue_type="human skeletal muscle"
                     /clone_lib="451 (synonym: hlcc1). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
     polyA_site      4538
ORIGIN

Alignment Scores:
Pred. No.:              5.66e-174       Length:         4560
Score:                  1979.00         Matches:        365
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0

DB:                     9               Gaps:           0

US-09-774-954-2 (1-365) x HSM803330 (1-4560)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         51 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 110

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        111 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 170

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        171 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 230

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        231 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 290

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        291 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 350

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        351 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 410

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        411 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 470

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        471 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 530

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        531 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 590

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        591 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 650

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        651 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 710

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        711 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 770

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        771 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 830

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        831 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 890

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        891 CAGTCGGTCTACGTTGCTACCGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 950

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        951 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1010

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1011 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1070

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1071 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1130

Qy        361 LeuArgAspGluPhe 365
                 |||||||||||||||
Db       1131 CTGCGGGACGAGTTC 1145

RESULT 10
D80002
LOCUS       D80002                  5189 bp    mRNA    linear   PRI 10-JAN-2003
DEFINITION  Homo sapiens KIAA0180 mRNA, complete cds.
ACCESSION   D80002
VERSION     D80002.2  GI:20521837
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nagase,T., Seki,N., Ishikawa,K., Tanaka,A. and Nomura,N.
  TITLE     Prediction of the coding sequences of unidentified human genes. V.
            The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
            analysis of cDNA clones from human cell line KG-1
  JOURNAL   DNA Res. 3 (1), 17-24 (1996)
  MEDLINE   96281124
   PUBMED   8724849
REFERENCE   2
  AUTHORS   Chiang,P.W., Wang,S., Smithivas,P., Song,W.J., Ramamoorthy,S.,
            Hillman,J., Puett,S., Van Keuren,M.L., Crombez,E., Kumar,A.,
            Glover,T.W., Miller,D.E., Tsai,C.H., Blackburn,C.C., Chen,X.N.,
            Sun,Z., Cheng,J.F., Korenberg,J.R. and Kurnit,D.M.
  TITLE     Identification and analysis of the human and murine putative
            chromatin structure regulator SUPT6H and Supt6h
  JOURNAL   Genomics 34 (3), 328-333 (1996)
  MEDLINE   96374824
   PUBMED   8786132
REFERENCE   3  (bases 1 to 5189)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1995) Osamu Ohara, Kazusa DNA Research Institute;
            1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913)
COMMENT     On May 9, 2002 this sequence version replaced gi:1136419.
FEATURES             Location/Qualifiers
     source          1. .5189
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="ha02567s1"
                     /sex="male"
                     /cell_line="KG-1"
                     /cell_type="myeloblast"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
                     /note="This sequence was obtained by subcloning of the DNA
                     fragments derived from two cDNA clones (1 - 1066 was
                     derived from pf00933 and 1067 - 5189 was derived from
                     ha02567)."
     gene            1. .5189
                     /gene="KIAA0180"
     CDS             <1. .1213
                     /gene="KIAA0180"
                     /citation=[2]
                     /codon_start=2
                     /protein_id="BAA11497.2"
                     /db_xref="GI:20521838"
                     /translation="PPRLCAAAGSGSRADMGAAAWARPLSVSFLLLLLPLPGMPAGSW
                     DPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSY
                     QKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEG
                     NPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE
                     EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHF
                     MASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPEL
                     QQLFKGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFG
                     MDRPPKLRDEF"
ORIGIN

Alignment Scores:
Pred. No.:                6.7e-174        Length:          5189
Score:                    1979.00         Matches:         365
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       9               Gaps:            0

US-09-774-954-2 (1-365) x D80002 (1-5189)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 175

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 235

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 295

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        296 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 355

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        356 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 415

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        416 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 475

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        476 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 535

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        536 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 595

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        596 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 655

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        656 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 715

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 775

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        776 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 835

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 895

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        896 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 955

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1015
```

```
Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1016 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1075

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1135

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1136 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1195

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1196 CTGCGGGACGAGTTC 1210
```

```
RESULT 11
AF375884
LOCUS       AF375884                5249 bp    mRNA    linear   PRI 23-OCT-2001
DEFINITION  Homo sapiens protein o-fucosyltransferase (POFUT1) mRNA, complete
            cds.
ACCESSION   AF375884
VERSION     AF375884.1  GI:15825113
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5249)
  AUTHORS   Wang,Y., Shao,L., Shi,S., Harris,R.J., Spellman,M.W., Stanley,P.
            and Haltiwanger,R.S.
  TITLE     Modification of epidermal growth factor-like repeats with O-fucose.
            Molecular cloning and expression of a novel GDP-fucose protein
            O-fucosyltransferase
  JOURNAL   J. Biol. Chem. 276 (43), 40338-40345 (2001)
  MEDLINE   21523965
   PUBMED   11524432
REFERENCE   2  (bases 1 to 5249)
  AUTHORS   Shi,S., Stanley,P., Wang,Y., Shao,L., Harris,R.J., Spellman,M.W.
            and Haltiwanger,R.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2001) Cell Biology, Albert Einstein College of
            Medicine, 1300 Morris Park Ave., New York, NY 10461, USA
FEATURES             Location/Qualifiers
     source          1. .5249
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="between PLAGL2 and KIF3B"
                     /tissue_type="heart"
     gene            1. .5249
                     /gene="POFUT1"
     CDS             50. .1216
                     /gene="POFUT1"
                     /function="catalyzes transfer of fucose from GDP-fucose to
                     serine or threonine on epidermal growth factor-like
                     repeats"
                     /note="O-FucT-1; glycosyltransferase"
                     /codon_start=1
                     /product="protein o-fucosyltransferase"
                     /protein_id="AAL09576.1"
                     /db_xref="GI:15825114"
                     /translation="MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRF
                     GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRV
                     ISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFN
                     KSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEM
                     VKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA
                     PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKP
                     EVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF"
ORIGIN

Alignment Scores:
```

```
Pred. No.:              6.8e-174        Length:         5249
Score:                  1979.00         Matches:        365
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-774-954-2 (1-365) x AF375884 (1-5249)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 598

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 658

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 718

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 778

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 838

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 898

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 958

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1018

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1078
```

```
Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1138

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1198

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1199 CTGCGGGACGAGTTC 1213

RESULT 12
AX780153
LOCUS       AX780153               5266 bp    DNA     linear   PAT 14-JUL-2003
DEFINITION  Sequence 2310 from Patent WO03039443.
ACCESSION   AX780153
VERSION     AX780153.1  GI:32697147
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Haferlach,T., Schoch,C., Kern,W., Kohlmann,A., Schnittger,S.,
            Dugas,M., Eils,R., Brors,B. and Mergenthaler,S.
  TITLE     Novel genetic markers for leukemias
  JOURNAL   Patent: WO 03039443-A 2310 15-MAY-2003;
            Deutsches Krebsforschungszentrum (DE) ;
            Ludwig-Maximilian-Universitaet Muenchen (DE) ; Haferlach, Torsten,
            PD Dr. Dr. (DE) ; Schoch, Claudia (DE) ; Kern, Wolfgang (DE)
FEATURES             Location/Qualifiers
     source          1. .5266
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                6.82e-174      Length:          5266
Score:                    1979.00        Matches:         365
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       6              Gaps:            0

US-09-774-954-2 (1-365) x AX780153 (1-5266)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 191

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 251

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 311

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 371

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 431

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 491

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 551

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 611

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 671

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 731

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        732 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 791

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        792 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 851

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 911

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 971

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1031

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1091

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1092 CTCGGNCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1151

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1211

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1212 CTGCGGGACGAGTTC 1226

RESULT 13
BD103190
LOCUS       BD103190               11284 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103190
VERSION     BD103190.1  GI:22648764
KEYWORDS    JP 2001527389-A/3.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 11284)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 3 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/3
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
```

```
          CC   Strandedness: Single;
          CC   Topology: Linear;
          CC   O-fucosyltransferase
          FH   Key             Location/Qualifiers
          FT   source          1. .11284
          FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .11284
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

Alignment Scores:
Pred. No.:              1.84e-173      Length:        11284
Score:                  1979.00        Matches:       365
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     6              Gaps:          0

US-09-774-954-2 (1-365) x BD103190 (1-11284)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4416 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 4475

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4476 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 4535

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4536 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 4595

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4596 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 4655

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4656 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 4715

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnProProValLeuGlu 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4716 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 4775

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4776 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 4835

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4836 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 4895

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4896 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 4955

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       4956 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 5015

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5016 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 5075

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5076 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 5135

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5136 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 5195

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5196 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 5255

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5256 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 5315

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       5316 CTGCGGGACGAGTTC 5330

RESULT 14
AY344580
LOCUS       AY344580               1176 bp    mRNA    linear   MAM 03-AUG-2003
DEFINITION  Bos taurus protein O-fucosyltransferase 1a (pofut1) mRNA, complete
            cds.
ACCESSION   AY344580
VERSION     AY344580.1  GI:33303529
KEYWORDS    .
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 1176)
  AUTHORS   Loriol,C., Germot,A., Dupuy,F. and Maftah,A.
  TITLE     Genomic organization and expression profile of O-fucosyltransferase
            genes, pofut1 and pofut2, in Bos taurus
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1176)
  AUTHORS   Loriol,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUL-2003) EA 3176 Glycobiologie et Biotechnologie,
            Faculte des Sciences et Techniques, 123, Avenue A. Thomas, Limoges
            87060, France
FEATURES             Location/Qualifiers
     source          1. .1176
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
     gene            1. .1176
                     /gene="pofut1"
     CDS             1. .1176
                     /gene="pofut1"
                     /note="O-FucT-1a"
                     /codon_start=1
                     /product="protein O-fucosyltransferase 1a"
                     /protein_id="AAQ02332.1"
                     /db_xref="GI:33303530"
                     /translation="MGAAAWAPSLLPPRVSLLLLLLPLPGLPVGSWDPAGYLLYCPCM
                     GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNVHVSYQKYFKLEPLQAY
                     HRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHV
                     SFNKSELFAGISFSASYKDQWIQRFSPEEHPVLALPGAPAQFPVLEEHRPLQKYMVWS
                     DEMVRTGEAQIHAHLIRPYVGIHLRIGSDWKNACAMLKDGTAGAHFMASPQCVGYSRH
                     TTAPLTMTMCLPDLKEIRRALKLWVTALNAQSVYIATDSESYLPEIQQLFKGKVKVVS
                     LKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLHGRPSSFFGMDRPPQLRDEF"
ORIGIN
```

```
Alignment Scores:
Pred. No.:              6.57e-167       Length:         1176
Score:                  1895.00         Matches:        341
Percent Similarity:     98.08%          Conservative:   17
Best Local Similarity:  93.42%          Mismatches:     7
Query Match:            95.76%          Indels:         0
DB:                     4               Gaps:           0

US-09-774-954-2 (1-365) x AY344580 (1-1176)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              :::|||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 CTGCCCGTGGGCTCCTGGGATCCGGCCGGTTACCTGCTCTACTGCCCATGCATGGGGCGT 138

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 TTTGGGAACCAGGCTGATCATTTCTTGGGCTCCCTGGCATTTGCAAAGCTGTTGAACCGC 198

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        199 ACCCTGGCTGTACCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACGTC 258

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCCTACCATCGAGTCATC 318

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 AGCCTGGAGGACTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 378

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGTCCTGATAAGAAGACATGCCCCATGAAG 438

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 GAAGGAAATCCCTTTGGCCCATTTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCCGAG 498

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||   |||||||||||||||||||||||:::::|||||   ||||||||||||
Db        499 CTTTTTGCAGGCATTTCCTTCAGTGCCTCCTACAAAGACCAGTGGATCCAGAGATTTTCC 558

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 CCAGAGGAACATCCAGTGCTTGCCCTGCCTGGTGCCCCTGCCCAGTTCCCTGTCCTGGAG 618

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        619 GAGCACAGGCCCCTCCAGAAGTACATGGTGTGGTCAGACGAGATGGTGAGGACAGGGGAG 678

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              |||||||||||||||||||||:::|||||||||||||||||||||||||||||||||||
Db        679 GCCCAGATCCACGCCCACCTCATCCGGCCCTACGTGGGCATTCACCTGCGCATTGGCTCT 738

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        739 GACTGGAAAAACGCGTGTGCCATGCTGAAAGATGGAACCGCTGGCGCCCACTTCATGGCC 798

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              |||||||||||||||||||||||||   |||   ||||||||||||||||||||||||
Db        799 TCCCCACAATGCGTGGGCTACAGCCGCCACACCACCGCCCCGCTTACCATGACCATGTGC 858

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              |||||||||||||||||||:::|||||:::||||||||||||   :::|||:::||||
Db        859 CTCCCTGACCTGAAGGAAATTAGGCGCGCTCTGAAGCTCTGGGTGACAGCGCTGAACGCC 918

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||:::|||||||||||||||||||:::||||||:::||||||||||||
Db        919 CAATCGGTCTACATTGCCACCGACTCTGAGAGTTACCTGCCCGAGATCCAGCAGCTCTTC 978

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db        979 AAAGGGAAGGTGAAGGTGGTGAGTCTGAAGCCTGAGGTGGCCCAGATCGACCTGTACATC 1038

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1039 CTCGGCCAAGCTGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTCGTAAAG 1098

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              |||||||||||||||   |||||||||||||||||||||||||||||||||||||:::
Db       1099 CGGGAACGGGACCTCCATGGGAGACCATCCTCCTTCTTTGGCATGGACAGGCCCCCTCAG 1158

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1159 CTGCGGGATGAATTC 1173

RESULT 15
AJ781503
LOCUS       AJ781503               1176 bp    mRNA    linear   MAM 07-JUL-2004
DEFINITION  Bos taurus mRNA for protein-O-fucosyltransferase 1 (fut12 gene).
ACCESSION   AJ781503
VERSION     AJ781503.1  GI:50057075
KEYWORDS    fut12 gene; pofut1 gene; protein-O-fucosyltransferase 1.
SOURCE      Bos taurus (cow)
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1
  AUTHORS   Martinez-Duncker,I., Mollicone,R., Candelier,J.J., Breton,C. and
            Oriol,R.
  TITLE     A new superfamily of protein-O-fucosyltransferases,
            alpha2-fucosyltransferases, and alpha6-fucosyltransferases:
            phylogeny and identification of conserved peptide motifs
  JOURNAL   Glycobiology 13 (12), 1C-5C (2003)
   PUBMED   12966037
REFERENCE   2
  AUTHORS   Martinez-Duncker,I., Oriol,R. and Mollicone,R.
  TITLE     Phylogeny of fucosyltransferases
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1176)
  AUTHORS   Oriol,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Oriol R., U504, Inserm, 16 Av. Paul
            Vaillant-Couturier, 94807, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1176
                     /organism="Bos taurus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9913"
     gene            1. .1176
                     /gene="fut12"
     CDS             1. .1176
                     /gene="fut12"
                     /function="adds fucose to Ser or Thr"
                     /note="pofut1"
                     /codon_start=1
                     /product="protein-O-fucosyltransferase 1"
                     /protein_id="CAH03714.1"
                     /db_xref="GI:50057076"
                     /translation="MGAAAWAPSLLPPRVSLLLLLPLPGLPVGSWNPAGYLLFCPCM
                     GRFGNQADHFLGSLAFAPLMNRTLAVPPWIEYQHHKPPFTNVHVSYQKYFKLEPLQAY
                     HRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHV
                     SFNKSELFAGISFSASYKDQWIQRFSPEEHPVLALPGAPAQFPVLEEHRPLQKYMVWS
                     DEMVRTGEAQIHAHLIRPYVGIHLRIGSDWKNACAMLKDGTAGAHFMASPQCVGYSRH
                     TTAPLTMTMCLPDLKEIRRALKLWVTALNAQSVYIATDSESYLPEIQQLFKGKVKVVS
                     LKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLHGRPSSFFGMDRPPQLRDEF"
ORIGIN

Alignment Scores:
Pred. No.:              2.52e-165       Length:         1176
Score:                  1878.00         Matches:        337
Percent Similarity:     97.81%          Conservative:   20
Best Local Similarity:  92.33%          Mismatches:     8
Query Match:            94.90%          Indels:         0
DB:                     4               Gaps:           0
```

```
US-09-774-954-2 (1-365) x AJ781503 (1-1176)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              :::|||   ||||||||:::||||||||||||||||:::||||||||||||||||||||
Db         79 CTGCCCGTGGGCTCCTGGAATCCGGCCGGTTACCTGCTTTTCTGCCCATGCATGGGGCGC 138

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||   |||:::||||||
Db        139 TTTGGGAACCAGGCTGATCATTTCTTGGGCTCCCTGGCATTTGCACCTCTGATGAACCGC 198

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        199 ACCCTGGCTGTACCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACGTC 258

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        259 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCCTACCATCGAGTCATC 318

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        319 AGCCTGGAGGACTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 378

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        379 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGTCCTGATAAGAAGACATGCCCCATGAAG 438

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        439 GAAGGAAATCCCTTTGGCCCATTTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCCGAG 498

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||   ||||||||||||||||||||||:::::||||||   ||||||||||||
Db        499 CTTTTGCAGGCATTTCCTTCAGTGCCTCCTACAAAGACCAGTGGATCCAGAGATTTTCC 558

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        559 CCAGAGGAACATCCAGTGCTTGCCCTGCCTGGTGCCCCTGCCCAGTTCCCTGTCCTGGAG 618

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db        619 GAGCACAGGCCCCTCCAGAAGTACATGGTGTGGTCAGACGAGATGGTGAGGACAGGGGAG 678

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db        679 GCCCAGATCCACGCCCACCTCATCCGGCCCTACGTGGGCATTCACCTGCGCATTGGCTCT 738

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||:::||||||||||||
Db        739 GACTGGAAAAACGCGTGTGCCATGCTGAAAGATGGAACCGCTGGCGCCCACTTCATGGCC 798

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||   |||   ||||||||||||||||||||||
Db        799 TCCCCACAATGCGTGGGCTACAGCCGCCACACCACCGCCCCGCTTACCATGACCATGTGC 858

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||:::|||||||:::||||||||||||   :::|||:::|||
Db        859 CTCCCTGACCTGAAGGAAATTAGGCGCGCTCTGAAGCTCTGGGTGACAGCGCTGAACGCC 918

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||:::|||||||||||||||||||:::||||||:::||||||||||||||
Db        919 CAATCGGTCTACATTGCCACCGACTCTGAGAGTTACCTGCCCGAGATCCAGCAGCTCTTC 978

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db        979 AAAGGGAAGGTGAAGGTGGTGAGTCTGAAGCCTGAGGTGGCCCAGATCGACCTGTACATC 1038

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1039 CTCGGCCAAGCTGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTCGTAAAG 1098

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||||:::

Db       1099 CGGGAACGGGACCTCCATGGGAGACCATCCTCCTTCTTTGGCATGGACAGGCCCCCTCAG 1158

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1159 CTGCGGGATGAATTC 1173


Search completed: October 26, 2005, 12:55:24
Job time : 8300.58 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:22:39 ; Search time 214.353 Seconds
(without alignments)
710.936 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....QGRPSSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1862994 seqs, 417510619 residues

Total number of hits satisfying chosen parameters: 1862994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1979	100.0	365	11	US-09-774-954-2	Sequence 2, Appli
2	1979	100.0	388	14	US-10-301-822-169	Sequence 169, App
3	1979	100.0	397	11	US-09-774-954-6	Sequence 6, Appli
4	1844	93.2	343	11	US-09-774-954-17	Sequence 17, Appl
5	831	42.0	402	20	US-11-097-143-18162	Sequence 18162, A
6	809	40.9	165	11	US-09-833-245-366	Sequence 366, App
7	541.5	27.4	474	11	US-09-774-954-8	Sequence 8, Appli
8	353	17.8	61	11	US-09-774-954-9	Sequence 9, Appli
9	295	14.9	61	11	US-09-774-954-3	Sequence 3, Appli
10	164.5	8.3	328	16	US-10-471-450-26	Sequence 26, Appl
11	159.5	8.1	490	20	US-11-097-143-36192	Sequence 36192, A
12	142	7.2	429	16	US-10-471-450-8	Sequence 8, Appli
13	120	6.1	150	18	US-10-820-474A-102	Sequence 102, App
14	117	5.9	543	15	US-10-424-599-145078	Sequence 145078,
15	106.5	5.4	441	16	US-10-437-963-162121	Sequence 162121,
16	103	5.2	561	16	US-10-425-115-328476	Sequence 328476,
17	103	5.2	581	15	US-10-425-114-65103	Sequence 65103, A
18	102.5	5.2	260	14	US-10-128-714-3312	Sequence 3312, Ap
19	101	5.1	28	11	US-09-774-954-14	Sequence 14, Appl
20	101	5.1	538	16	US-10-437-963-113919	Sequence 113919,
21	100	5.1	207	15	US-10-309-290-120	Sequence 120, App
22	100	5.1	207	15	US-10-309-290-122	Sequence 122, App
23	98.5	5.0	446	15	US-10-425-114-68801	Sequence 68801, A
24	95.5	4.8	442	16	US-10-425-115-237127	Sequence 237127,
25	95	4.8	385	14	US-10-156-761-13301	Sequence 13301, A
26	94.5	4.8	4773	15	US-10-287-226-322	Sequence 322, App
27	93	4.7	713	16	US-10-437-963-170855	Sequence 170855,
28	93	4.7	1374	16	US-10-437-963-132782	Sequence 132782,
29	92.5	4.7	1251	15	US-10-282-122A-63230	Sequence 63230, A
30	92	4.6	348	15	US-10-282-122A-53200	Sequence 53200, A
31	92	4.6	650	15	US-10-282-122A-73270	Sequence 73270, A
32	91	4.6	1381	14	US-10-032-585-7784	Sequence 7784, Ap
33	91	4.6	1381	16	US-10-451-467A-426	Sequence 426, App
34	90.5	4.6	353	15	US-10-309-290-118	Sequence 118, App
35	90.5	4.6	455	18	US-10-450-763-55399	Sequence 55399, A
36	90.5	4.6	568	15	US-10-282-122A-71715	Sequence 71715, A
37	90	4.5	1205	16	US-10-437-963-161150	Sequence 161150,
38	90	4.5	1315	18	US-10-450-763-40320	Sequence 40320, A
39	89.5	4.5	218	17	US-10-732-923-10961	Sequence 10961, A
40	89	4.5	511	16	US-10-425-115-324710	Sequence 324710,
41	89	4.5	518	15	US-10-425-114-53636	Sequence 53636, A
42	89	4.5	555	16	US-10-425-115-357766	Sequence 357766,
43	89	4.5	581	15	US-10-425-114-59580	Sequence 59580, A
44	88.5	4.5	314	16	US-10-425-115-283759	Sequence 283759,
45	88.5	4.5	991	16	US-10-437-963-181268	Sequence 181268,

ALIGNMENTS

RESULT 1
US-09-774-954-2
; Sequence 2, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:

```
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 2:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 365 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-774-954-2

  Query Match            100.0%;  Score 1979;  DB 11;  Length 365;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-195;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360

Qy        361 LRDEF 365
              |||||
Db        361 LRDEF 365


RESULT 2
US-10-301-822-169
; Sequence 169, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
;  APPLICANT:  Burgart, Lawrence J.
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
;  CURRENT APPLICATION NUMBER: US/10/301,822
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 169
;   LENGTH: 388
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-10-301-822-169

  Query Match            100.0%;  Score 1979;  DB 14;  Length 388;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-195;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 83

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         84 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 143

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 203

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 263

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 323

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 383

Qy        361 LRDEF 365
              |||||
Db        384 LRDEF 388


RESULT 3
US-09-774-954-6
; Sequence 6, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 6:
```

```
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 397 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-774-954-6

  Query Match             100.0%;  Score 1979;  DB 11;  Length 397;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-195;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         33 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 92

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 152

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 212

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 272

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 332

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 392

Qy        361 LRDEF 365
              |||||
Db        393 LRDEF 397


RESULT 4
US-09-774-954-17
; Sequence 17, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 17:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 343 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-774-954-17

  Query Match             93.2%;  Score 1844;  DB 11;  Length 343;
  Best Local Similarity   100.0%;  Pred. No. 9.6e-182;
  Matches  343;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         23 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRVISL 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRVISL 60

Qy         83 EDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELF 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELF 120

Qy        143 TGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQ 202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQ 180

Qy        203 IHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLP 262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLP 240

Qy        263 DLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILG 322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 DLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILG 300

Qy        323 QADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 365
              |||||||||||||||||||||||||||||||||||||||||||
Db        301 QADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 343


RESULT 5
US-11-097-143-18162
; Sequence 18162, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18162
;   LENGTH: 402
```

```
;   TYPE: PRT
;   ORGANISM: DROSOPHILA
US-11-097-143-18162

  Query Match             42.0%;  Score 831;  DB 20;  Length 402;
  Best Local Similarity   44.7%;  Pred. No. 9e-77;
  Matches  167;  Conservative   62;  Mismatches  119;  Indels   26;  Gaps    9;

Qy          7 DPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQK 66
              || ||| |||||||||||||||||||||||| ||||| :|||:||:  :    :  | :
Db         28 DPNGYLTYCPCMGRFGNQADHFLGSLAFAKALNRTLILPPWVEYR--RGELRSRQVPFNT 85

Qy         67 YFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFE-----VAAQRSPDKKTCPMKE 121
              ||::|||: |||||:: |||  ||   ||  :||::|::       :  |||  |  |:
Db         86 YFEVEPLKEYHRVITMADFMWHLADDIWPESERVSFCYKERYSLQQEKNDPDKPNCHAKD 145

Qy        122 GNPFGPFWDQFHVSFNKSELFTGISFSASYREQ---WSQRFSPKEHPVLALPGAPAQFPV 178
              ||||||||| ||: | :|| :  : |   :  :   |  ::  : :||||  |||| |||
Db        146 GNPFGPFWDTFHIDFVRSEFYAPLHFDVHHSNEAAKWQTKYPAESYPVLAFTGAPASFPV 205

Qy        179 LEEHRPLQKYMVWSDEMVKTGEAQIHAHLVR-PYVGIHLRIGSDWKNACAMLKDGTAGSH 237
                |:  ||:|: ||    :  :  |   | |  ::|||| | ||  ||  :||     |
Db        206 QLENCKLQRYLQWSQRYREASKDFIREQLPRGAFLGIHLRNGIDWVRACEHVKD---SQH 262

Qy        238 FMASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSL-------DAQSVYVATDSE 290
              |||||:||     | |   :|:|  : | | :|  ::::       : :||:||:|| 
Db        263 LFASPQCLGYKNERGA-LYPELCMPSKEAIIRQLKRTIKNVRQTQPDNEIKSVFVASDSN 321

Qy        291 SYVPELQ-QLFKGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRP 349
              : ||   | :  : |  |  :   :|| ||||::|||||:||::|| |||||: | |
Db        322 HMIGELNTALSRMGISVHKLPEDDPYLDLAILGQSNHFIGNCISSYSAFEKRERDVHGFP 381

Qy        350 SSFFGMDRPPKLRD 363
              | |:|     || :|
Db        382 SYFWGF---PKEKD 392


RESULT 6
US-09-833-245-366
; Sequence 366, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:
;  APPLICANT: Human Genome Sciences, Inc.
;  TITLE OF INVENTION: Albumin Fusion Proteins
;  FILE REFERENCE: PF546PCT
;  CURRENT APPLICATION NUMBER: US/09/833,245
;  CURRENT FILING DATE:  2001-04-12
;  PRIOR APPLICATION NUMBER: 60/229, 358
;  PRIOR FILING DATE: 2000-04-12
;  PRIOR APPLICATION NUMBER: 60/256, 931
;  PRIOR FILING DATE: 2000-12-21
;  PRIOR APPLICATION NUMBER: 60/199, 384
;  PRIOR FILING DATE: 2000-04-25
;  NUMBER OF SEQ ID NOS: 2267
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 366
;   LENGTH: 165
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (2)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (4)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (5)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (148)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-833-245-366

  Query Match             40.9%;  Score 809;  DB 11;  Length 165;
  Best Local Similarity   96.9%;  Pred. No. 4.7e-75;
  Matches  155;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        206 HLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLK 265
              |  | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          6 HFARTYPGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLK 65

Qy        266 EIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILGQAD 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 EIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILGQAD 125

Qy        326 HFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 365
              |||||||||||||||||| ||||||||||||||||||
Db        126 HFIGNCVSSFTAFVKRERDLQGXPSSFFGMDRPPKLRDEF 165


RESULT 7
US-09-774-954-8
; Sequence 8, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 8:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 474 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-774-954-8

  Query Match             27.4%;  Score 541.5;  DB 11;  Length 474;
  Best Local Similarity   36.2%;  Pred. No. 1.1e-46;
  Matches  125;  Conservative   60;  Mismatches  107;  Indels   53;  Gaps   11;

Qy         19 GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHR 78
              |||||| | ||| ||||| |:||| :| :||::|   | | : : ::  |::  :  | |
Db        166 GRFGNQVDQFLGVLAFAKALDRTLVLPNFIEFKH---PETKM-IPFEFLFQVGTVAKYTR 221

Qy         79 VISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNK 138
              |:::::| :|: |||:    | |  :: :|:       |  |||||||:|||  |||
Db        222 VVTMQEFTKKIMPTHFVGTPRQA-IYDKSAEPG-----CHSKEGNPFGPYWDQIDVSFVG 275
```

```
Qy        139 SELFTGI--SFSASY---REQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSD 193
              | |  |   |  :    |::| ::|  :|:||||   ||| ||   :   :|||: ||
Db        276 DEYFGDIPGGFDLNQMGSRKKWLEKFPSEEYPVLAFSSAPAPFPSKGKVWSIQKYLRWSS 335

Qy        194 EMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA 253
               : :  :  | |:| :|:| :|||  :||   |  : | |       || ||:|
Db        336 RITEQAKKFISANLAKPFVAVHLRNDADWVRVCEHI-DTTTNRPLFASEQCLGEGHHLGT 394

Qy        254 PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEV 313
              ||  :| |  ::|                              |:|     ::    :|:
Db        395 -LTKEICSPSKQQI--------------------------LEQ-----IEAHRQEPDD 420

Qy        314 AQVDLYILGQADHFIGNCVSSFTAFVKRERDLQG---RPSSFFGM 355
                | |:|:|| |:|||||:|:  ||||||  |   |||:|||:
Db        421 MYTSLAIMGRADLFVGNCVSTFSHIVKRERDHAGQSPRPSAFFGI 465


RESULT 8
US-09-774-954-9
; Sequence 9, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 9:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 61 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-774-954-9

  Query Match             17.8%;  Score 353;  DB 11;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-28;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 9
US-09-774-954-3
; Sequence 3, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 3:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 61 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-774-954-3

  Query Match             14.9%;  Score 295;  DB 11;  Length 61;
  Best Local Similarity   91.5%;  Pred. No. 1.8e-22;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||
Db          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 10
US-10-471-450-26
; Sequence 26, Application US/10471450
; Publication No. US20040152877A1
; GENERAL INFORMATION:
;  APPLICANT: JACKSON, Jennifer L.; BAUGHN, Mariah R.;
;  APPLICANT:  SWARNAKAR, Anita; YUE, Henry;
;  APPLICANT:  ELLIOTT, Vicki S.; BURFORD, Neil;
;  APPLICANT:  DING, Li; TANG, Y. Tom;
;  APPLICANT:  LEE, Soo Yeun; AZIMZAI, Yalda;
;  APPLICANT:  CHAWLA, Narinder K.;GIETZEN, Kimberly J.;
;  APPLICANT:  GRIFFIN, Jennifer A.; LAL, Preeti G.;
;  APPLICANT:  YANG, Junming; BOROWSKY, Mark L.;
;  APPLICANT:  RICHARDSON, Thomas W.; YUE, Huibin;
;  APPLICANT:  BECHA, Shanya; FORSYTHE, Ian J.;
;  APPLICANT:  JONES, Karen Anne; WARREN, Bridget;
;  APPLICANT:  THANGAVELU, Kavitha; HONCHELL, Cynthia;
;  APPLICANT:  JOLLEY, Helen E.;HAFALIA, April J.A.; Huijun Z. Ring
```

```
;  TITLE OF INVENTION: NUCLEIC ACID ASSOCIATED PROTEINS
;  FILE REFERENCE: PF-0917 USN
;  CURRENT APPLICATION NUMBER: US/10/471,450
;  CURRENT FILING DATE:  2003-09-10
;  PRIOR APPLICATION NUMBER: PCT/US02/07869
;  PRIOR FILING DATE: 2002-03-14
;  PRIOR APPLICATION NUMBER: US 60/276,857
;  PRIOR FILING DATE: 2001-03-16
;  PRIOR APPLICATION NUMBER: US 60/285,489
;  PRIOR FILING DATE: 2001-04-19
;  PRIOR APPLICATION NUMBER: US 60/285,556
;  PRIOR FILING DATE: 2001-04-19
;  PRIOR APPLICATION NUMBER: US 60/288,700
;  PRIOR FILING DATE: 2001-05-04
;  PRIOR APPLICATION NUMBER: US 60/288,646
;  PRIOR FILING DATE: 2001-05-04
;  PRIOR APPLICATION NUMBER: US 60/290,510
;  PRIOR FILING DATE: 2001-05-10
;  PRIOR APPLICATION NUMBER: US 60/290,369
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/332,426
;  PRIOR FILING DATE: 2001-11-16
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: PERL Program
; SEQ ID NO 26
;   LENGTH: 328
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 7503717CD1
US-10-471-450-26

  Query Match              8.3%;  Score 164.5;  DB 16;  Length 328;
  Best Local Similarity   21.7%;  Pred. No. 6.7e-08;
  Matches   85;  Conservative   41;  Mismatches  115;  Indels  151;  Gaps   15;

Qy          3 AGSWDPAG------------------------YLLY--CPCMGRFGNQADHFLGSLAFAK 36
                | || ||                         ||||   |  |  |   : |  ::   :  |
Db         13 AVSWPPASASGQEFWPGQSAADILSGAASRRRYLLYDVNPPEG-FNLRRDVYIRIASLLK 71

Qy         37 LLNRT----LAVPPWIEYQH-HKPPFTNLHVSYQKYFKLEPLQAYHRVISLEDFMEKLAP 91
                | :|      |  :|||    |    |    : :  : ::| |  |       ||  | |: :
Db         72 TLLKTEEWVLVLPPWGRLYHWQSPDIHQVRIPWSEFFDLPSLNKNIPVIEYEQFIAESG- 130

Qy         92 THWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELFTGISFSASY 151
                                                   ||| ||  :|                 ||
Db        131 ---------------------------------GPFIDQVYV-------------LQSY 143

Qy        152 REQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMV-----WSDEMVKTGEAQIHAH 206
                | |  :         |    |                  : ||     ::       |    || |   |
Db        144 AEGWKE--GTWEEKV--------------DERPCIDQLLYFQEDWMKMKVKLGSA----- 182

Qy        207 LVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLKE 266
                |   ||:|:|||               ||   |                                 :| |:
Db        183 LGGPYLGVHLR-----------RKDFIWGHR--------------------QDVPSLEG 210

Qy        267 IQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPE-----------VAQ 315
                 | ::  :::       |:||||:      |   ::|  |     ::|   :|                  ||
Db        211 AVRKIRSLMKTHRLDKVFVATDAVR--KEYEELKKLLPEMVRFEPTWEELELYKDGGVAI 268

Qy        316 VDLYILGQADHFIGNCVSSFTAFVKRERDLQG 347
                :| :|    |  |||   ||:|:   :   ||:: |
Db        269 IDQWICAHARFFIGTSVSTFSFRIHEEREILG 300
```

```
RESULT 11
US-11-097-143-36192
; Sequence 36192, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 36192
;   LENGTH: 490
;   TYPE: PRT
;   ORGANISM: DROSOPHILA
US-11-097-143-36192

  Query Match              8.1%;  Score 159.5;  DB 20;  Length 490;
  Best Local Similarity   22.4%;  Pred. No. 4e-07;
  Matches   92;  Conservative   55;  Mismatches  159;  Indels  105;  Gaps   19;

Qy          9 AGYLLYCPCMGR-FGNQADHFLGSLAFAKLLNR-------TLAVPPWIEYQH-HKPPFTN 59
                | |:||    :     |   :  |  ::       |  :  |  |          | :|||     | |
Db         71 AVYILYDVNISEGFNLRRDVYIRMAVFVRRLQRRRRFRHVRLVLPPWPRLYHWHSQGLQQ 130

Qy         60 LHVSYQKYFKLEPLQAYHRVISLEDFM--EKLAPTHWPPEKRVAYCFEV----------- 106
                 :  :  :|  |   |:  |  |:   |:|:  ::|       |      |  :   |  :
Db        131 SGLPWSHFFDLASLRRYAPVLDYEEFLAEQRLFGNPGAPLVHVGHAFRLQHYEVMLEQGI 190

Qy        107 ---AAQRSPDKKTCPMKEGN-PFGPFWDQFHVSFNKSELFTGISFSAS-------YREQW 155
                   :|  ||    |   ||:    ||     |    :      :    |   :  |   |            ||
Db        191 FRDKFERVTDK---PCSEGSLSGGPLLQQAELRVGR---FHCVRFQGSAGLLEKLLREAI 244

Qy        156 SQRFSPKEH----PVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQIHAHLVRPY 211
                 :   :    |                  ||   |      || :|             | ||                |
Db        245 DEDTAGPEDVDDMRTYALLSAET---VLHDH--------WGDE---------HFWQARRS 284

Qy        212 VGIHLR---IGSDWKNACAMLKDGTA-------------------GSHFMASPQCVGYSR 249
                :      |      : :|::              | :|                         | :   |   :   : |
Db        285 MRFARRLEQVAADFRRQALDTTDASAGVQRPAMWELERPKRNAKGGDYLCAHLRRGDFVR 344

Qy        250 STAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESY-VPELQQLFKGKVKVVS 308
                |     |        | ||     :  ||   :|   :   :|::|||:   |  :  ||::||    : ::|
Db        345 SRDA------TTPTLKAAAQQVKQLLRGFNMTTVFLATDATPYELMELKELFY-RFRLVH 397

Qy        309 LKPE-----------VAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQG 347
                  ||                   || ||    :     | :|:|    |:||   :   ||:: |
Db        398 FAPESNVQRRELKDGGVAVVDQLVCAYARYFVGTYESTFTYRIYEEREILG 448
```

```
RESULT 12
US-10-471-450-8
; Sequence 8, Application US/10471450
; Publication No. US20040152877A1
; GENERAL INFORMATION:
;  APPLICANT: JACKSON, Jennifer L.; BAUGHN, Mariah R.;
;  APPLICANT:  SWARNAKAR, Anita; YUE, Henry;
;  APPLICANT:  ELLIOTT, Vicki S.; BURFORD, Neil;
;  APPLICANT:  DING, Li; TANG, Y. Tom;
;  APPLICANT:  LEE, Soo Yeun; AZIMZAI, Yalda;
```

```
;  APPLICANT:  CHAWLA, Narinder K.;GIETZEN, Kimberly J.;
;  APPLICANT:  GRIFFIN, Jennifer A.; LAL, Preeti G.;
;  APPLICANT:  YANG, Junming; BOROWSKY, Mark L.;
;  APPLICANT:  RICHARDSON, Thomas W.; YUE, Huibin;
;  APPLICANT:  BECHA, Shanya; FORSYTHE, Ian J.;
;  APPLICANT:  JONES, Karen Anne; WARREN, Bridget;
;  APPLICANT:  THANGAVELU, Kavitha; HONCHELL, Cynthia;
;  APPLICANT:  JOLLEY, Helen E.;HAFALIA, April J.A.; Huijun Z. Ring
;  TITLE OF INVENTION: NUCLEIC ACID ASSOCIATED PROTEINS
;  FILE REFERENCE: PF-0917 USN
;  CURRENT APPLICATION NUMBER: US/10/471,450
;  CURRENT FILING DATE:  2003-09-10
;  PRIOR APPLICATION NUMBER: PCT/US02/07869
;  PRIOR FILING DATE: 2002-03-14
;  PRIOR APPLICATION NUMBER: US 60/276,857
;  PRIOR FILING DATE: 2001-03-16
;  PRIOR APPLICATION NUMBER: US 60/285,489
;  PRIOR FILING DATE: 2001-04-19
;  PRIOR APPLICATION NUMBER: US 60/285,556
;  PRIOR FILING DATE: 2001-04-19
;  PRIOR APPLICATION NUMBER: US 60/288,700
;  PRIOR FILING DATE: 2001-05-04
;  PRIOR APPLICATION NUMBER: US 60/288,646
;  PRIOR FILING DATE: 2001-05-04
;  PRIOR APPLICATION NUMBER: US 60/290,510
;  PRIOR FILING DATE: 2001-05-10
;  PRIOR APPLICATION NUMBER: US 60/290,369
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/332,426
;  PRIOR FILING DATE: 2001-11-16
;  NUMBER OF SEQ ID NOS: 52
;  SOFTWARE: PERL Program
; SEQ ID NO 8
;  LENGTH: 429
;  TYPE: PRT
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  OTHER INFORMATION: Incyte ID No: 7493913CD1
US-10-471-450-8

  Query Match             7.2%;  Score 142;  DB 16;  Length 429;
  Best Local Similarity   22.0%;  Pred. No. 2.1e-05;
  Matches   99;  Conservative   46;  Mismatches  138;  Indels  168;  Gaps   19;

Qy          3 AGSWDPAG-----------------------YLLY--CPCMGRFGNQADHFLGSLAFAK 36
              | || ||                       ||||   |  | |  : | ::   :  |
Db         13 AVSWPPASASGQEFWPGQSAADILSGAASRRRYLLYDVNPPEG-FNLRRDVYIRIASLLK 71

Qy         37 LLNRT----LAVPPWIEYQH-HKPPFTNLHVSYQKYFKLEPLQAYHRVISLEDFMEKLAP 91
               | :|    | :|||    |   |    : : : ::| |  |      ||  | |   :|
Db         72 TLLKTEEWVLVLPPWGRLYHWQSPDIHQVRIPWSEFFDLPSLNKNIPVIEYEQF---IAE 128

Qy         92 THWPPEKRVAYCFEVAAQRSPDKKTCPMKEGN-----PFGPFWDQFHVSFNKSELFTG-- 144
              :  |   :| |  :  |:         |||          |  ||    | :| | : |
Db        129 SGGPFIDQV-YVLQSYAE--------GWKEGTWEEKVDERPCIDQLLYSQDKHEYYRGWF 179

Qy        145 ---------------ISFSASY----------------------------REQWSQRFS 160
                           :  |||                           :| |  | |
Db        180 WGYEETRGLNVSCLSVQGSASIVAPLLLRNTSARSVMLDRAENLLHDHYGGKEYWDTRRS 239

Qy        161 PKEHPVLALPGAPAQFPVLEEHR-------------PLQKYMVWSDEMVKTGEAQIHAHL 207
                      :  |      | :| |              | |:   |     || | |     |
Db        240 --------MVFARHLREVGDEFRSRHLNSTDDADRIPFQE--DWMKMKVKLGSA-----L 284

Qy        208 VRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLKEI 267
                ||:|:|||              ||   |                          :| |:
Db        285 GGPYLGVHLR-----------RKDFIWGHR-----------------'----QDVPSLEGA 312

Qy        268 QRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPE-----------VAQV 316
               | ::  :::     |:||||:     | ::| |   ::|  :|              || :
Db        313 VRKIRSLMKTHRLDKVFVATDAVR--KEYEELKKLLPEMVRFEPTWEELELYKDGGVAII 370

Qy        317 DLYILGQADHFIGNCVSSFTAFVKRERDLQG 347
              | :|   |  |||  ||:|:  :  ||:: |
Db        371 DQWICAHARFFIGTSVSTFSFRIHEEREILG 401


RESULT 13
US-10-820-474A-102
; Sequence 102, Application US/10820474A
; Publication No. US20050155089A1
; GENERAL INFORMATION:
;  APPLICANT: LAL, PREETI
;  APPLICANT:  TANG, Y. TOM
;  APPLICANT:  GORGONE, GINA A.
;  APPLICANT:  CORLEY, NEIL C.
;  APPLICANT:  GUEGLER, KARL J.
;  APPLICANT:  BAUGHN, MARIAH R.
;  APPLICANT:  AKERBLOM, INGRID E.
;  APPLICANT:  AU-YOUNG, JANICE
;  APPLICANT:  YUE, HENRY
;  APPLICANT:  PATTERSON, CHANDRA
;  APPLICANT:  REDDY, ROOPA
;  APPLICANT:  HILLMAN, JENNIFER L.
;  APPLICANT:  BANDMAN, OLGA
;  TITLE OF INVENTION: SIGNAL PEPTIDE-CONTAINING MOLECULES
;  FILE REFERENCE: 039386-1568
;  CURRENT APPLICATION NUMBER: US/10/820,474A
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: 09/720,533
;  PRIOR FILING DATE: 2001-03-20
;  PRIOR APPLICATION NUMBER: PCT/US99/14484
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/090,762
;  PRIOR FILING DATE: 1998-06-26
;  PRIOR APPLICATION NUMBER: 60/094,983
;  PRIOR FILING DATE: 1998-07-31
;  PRIOR APPLICATION NUMBER: 60/102,686
;  PRIOR FILING DATE: 1998-10-01
;  NUMBER OF SEQ ID NOS: 269
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 102
;  LENGTH: 150
;  TYPE: PRT
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  OTHER INFORMATION: Incyte Clone No: 1647884
US-10-820-474A-102

  Query Match             6.1%;  Score 120;  DB 18;  Length 150;
  Best Local Similarity   95.0%;  Pred. No. 0.00087;
  Matches   19;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGR 20
              ||||||||||||||||||:
Db         24 MPAGSWDPAGYLLYCPCMGK 43


RESULT 14
US-10-424-599-145078
; Sequence 145078, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 145078
```

```
;    LENGTH: 543
;    TYPE: PRT
;    ORGANISM: Glycine max
;    FEATURE:
;    OTHER INFORMATION: Clone ID: PAT_MRT3847_102023C.1.pep
US-10-424-599-145078

  Query Match              5.9%;  Score 117;  DB 15;  Length 543;
  Best Local Similarity   22.0%;  Pred. No. 0.012;
  Matches   89;  Conservative   50;  Mismatches  137;  Indels  128;  Gaps   23;

Qy          6 WDPA-GYLLYCPCMGRFGNQADHFL---GSLAFAKLLNRTLAVP-PWIEYQHHK------ 54
                |:|   |   |     |:    |   ::|  :      :  || ||||  | :|    ::||: :
Db        190 WNPKDGKFLLAICVS--GQMSNHLICLEKHIFFAALLNRVLVIPSSKVDYQYDRVVDIDH 247

Qy         55 ------------PPFTNL---HVSYQK---YFKLEPLQAYHRVISLED-FMEKLAPTHW 94
                          |:||    |:    |     ||   :|    |         |:|   ::||
Db        248 INKCLGKKVVVSFDVFSNLKKGHLHIDKFLCYFS-QPQPCY-----LDDERLKKLGALGL 301

Qy         95 PPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELFTGISFSASYREQ 154
                    |   |    |       | |  |||          :   | |        ||:     :    |   |  |       :
Db        302 TMSKPEAVWDE--DTRKPKKKTVQ----DVLGKF------SFDDDVMAIGDVFYAEVERE 349

Qy        155 WSQRFSPKEHPVLALPGAPAQF---PVLEEHRPLQKYMVWSDEMVKTGEAQIHAHLVRPY 211
              |   :            || |          ::| :| :                :  | :   |     | |  :
Db        350 WVMQ----------PGGPIAHKCKTLIEPNRLI---------LLTAQRFIQTFLGRNF 388

Qy        212 VGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAV 271
               : :|  |         ::              |    |         ||                |:    |:  :: |
Db        389 IALHFR-----RHGFLKFCNAKKPSCFYPIPQAAD-----------CI--LRVVEMA- 427

Qy        272 KLWVRSLDAQSVYVATD-SESYVPELQQL--FKGKVKVVSLKPE--------------- 312
                     .     ||  :|::|| :||       || |       |:     : ::|
Db        428 -------DAPIIYLSTDAAESETGLLQSLVVLNGRPVPLVIRPARNSAEKWDALLYRHNM 480

Qy        313 ------VAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPS 350
                    | :|    |      :    |||     |:||     :   |    |        |
Db        481 DGDSQVEAMLDKTICAMSSVFIGAPGSTFTEDILRLRKDWGSAS 524
```

RESULT 15

```
US-10-437-963-162121
; Sequence 162121, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 162121
;   LENGTH: 441
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_61241C.1.pep
US-10-437-963-162121

  Query Match              5.4%;  Score 106.5;  DB 16;  Length 441;
  Best Local Similarity   19.9%;  Pred. No. 0.1;
  Matches   77;  Conservative   63;  Mismatches  147;  Indels   99;  Gaps   17;

Qy          4 GSWDPAGYLLYCPCMGRF----------GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHH 53
                | :||    ||      |:|               ||     |    :      |:  |||:|  :|         :
```

```
Db         51 GFYDPTADLLPGIRRGKFIEVPQIIWGLNNQKIAFARACLTARFLNRSLLMPSLSASLFY 110

Qy         54 KPPFTNLHVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPD 113
              |         :::  |  |         |             : |:                   |:|    ||:  |   |
Db        111 KEVDLLRPITFDKVFDFTKFNA-----RCQGFV------------RLARYSEVSNQTKPF 153

Qy        114 KKTCPMKEGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGA- 172
              |      :::|:   |    |       |   :   :|              ||      |       :    |:  :  |
Db        154 K----LQKGS--GRRWT---VEKDLDQLL-------QYR-----RGEADDSEVIEIIGKH 192

Qy        173 PAQFPVLEEHRPLQKYMVWSD----------EMVKT----GEAQIHA----------HLV 208
              |   :|     :|  |:: |      |                 |:||       || | |              |:
Db        193 PFLWP---DHWPVKDYARIFDCLALVPEIETEVVKAISKIREAGIKARHEAGISHNKHVK 249

Qy        209 R-------PYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCL 261
                        ||: :|:||   ||     |      :     :          :| : : :    |
Db        250 DGTMNPPVPYIAVHMRIEKDWMIHCKKWEQRSNSKEICSSKEEIIHKVS----------- 298

Qy        262 PDLKEIQRAVKLWVRSLDA--QSVYVATDSESYVPELQQLFKGKVKVVSLKPEV--AQVD 317
                : :::|  | :::      |:    :      :    :         :      ::     |      : : :|  :     : :|
Db        299 -QITDLRRPVVVYLAVADSLLEDDSITSGWRVGMVAFEKKRLGVTDIYNRQPYLIKSAID 357

Qy        318 LYILGQADHFIGNCVSSFTAFVKRER 343
                 :    :||  |:||   |:|:     |      |
Db        358 FEVCARADVFVGNSFSTFSNLVVLSR 383
```

Search completed: October 25, 2005, 15:41:21
Job time : 216.353 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 03:15:37 ; Search time 1022.3 Seconds
(without alignments)
2113.573 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....QGRPSSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105431_6380/app_query.fasta_1.917
-DB=N_Geneseq -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954_@CGN_1_1_703_@runat_25102005_105431_6380 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1979	100.0	1514	2	AAV65632	Aav65632 Human hea
	2	1979	100.0	5218	13	ACN40746	Acn40746 Tumour-as
	3	1979	100.0	5266	10	ADF81754	Adf81754 Leukaemia
	4	1979	100.0	11284	2	AAV65633	Aav65633 Plasmid c
	5	1867.5	94.4	5230	4	AAK51510	Aak51510 Human pol
	6	1855.5	93.8	4850	4	AAK52494	Aak52494 Human pol
	7	1844	93.2	5009	2	AAV65634	Aav65634 First Eco
	8	831	42.0	1209	4	ABL13947	Abl13947 Drosophil
	9	811	41.0	610	13	ADQ56824	Adq56824 Novel can
	10	809	40.9	915	4	AAD07629	Aad07629 Human sec
c	11	796	40.2	3264	4	ABL13946	Abl13946 Drosophil
c	12	506.5	25.6	3793	4	ABL13894	Abl13894 Drosophil
	13	424	21.4	10331	4	AAK79915	Aak79915 Human imm
	14	424	21.4	10331	4	AAL03344	Aal03344 Human rep
	15	393	19.9	479	9	ACH32225	Ach32225 Human end
c	16	392	19.8	3567	4	ABL13952	Abl13952 Drosophil
	17	353	17.8	477	9	ACH35532	Ach35532 Human end
	18	219	11.1	1545	4	AAF23894	Aaf23894 Human sec
	19	169.5	8.6	2229	9	ADA21149	Ada21149 Human sec
	20	166.5	8.4	2151	9	ADA21151	Ada21151 Human sec
	21	164.5	8.3	1967	6	ABS73886	Abs73886 Human cDN
	22	164	8.3	1473	13	ADS96503	Ads96503 Drosophil
	23	164	8.3	1926	4	ABL25967	Abl25967 Drosophil
c	24	164	8.3	3926	4	ABL25966	Abl25966 Drosophil
	25	149.5	7.6	2374	9	ADA21150	Ada21150 Human sec
	26	142	7.2	2434	6	ABS73868	Abs73868 Human cDN
	27	120	6.1	760	3	AAZ98210	Aaz98210 Human sig
c	28	115.5	5.8	4104	4	ABL25964	Abl25964 Drosophil
	29	106	5.4	2365	9	ADA21178	Ada21178 Human sec
	30	105.5	5.3	4845	3	AAC74541	Aac74541 Human ORF
	31	104	5.3	2071	3	AAC52262	Aac52262 Arabidops
	32	104	5.3	97081	12	ADQ97980	Adq97980 Human can
	33	103.5	5.2	960	10	ADE95589	Ade95589 Human NOV
	34	103.5	5.2	960	10	ADE95587	Ade95587 Human NOV
	35	102.5	5.2	783	8	ABT19142	Abt19142 Aspergill
	36	102.5	5.2	907	8	ABT18548	Abt18548 Aspergill
	37	102.5	5.2	2162	4	AAK73291	Aak73291 Human imm
	38	102.5	5.2	2907	8	ABT17954	Abt17954 Aspergill
	39	102	5.2	2069	3	AAC54555	Aac54555 Arabidops
	40	102	5.2	29069	3	AAA81497	Aaa81497 N. mening
	41	102	5.2	110000	3	AAA81490_11	Continuation (12 o
	42	102	5.2	349980	3	AAF21609	Aaf21609 Neisseria
	43	101.5	5.1	2366	1	AAN91052	Aan91052 Phage DNA
	44	101.5	5.1	2381	2	AAQ05241	Aaq05241 Sequence
	45	101.5	5.1	2381	10	ADB57903	Adb57903 Toxicity-

ALIGNMENTS

RESULT 1
AAV65632
ID AAV65632 standard; DNA; 1514 BP.
XX
AC AAV65632;
XX
DT 16-DEC-1998 (first entry)
XX
DE Human heart O-fucosyltransferase encoding DNA.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_feature 1. .1100
FT /*tag= b
FT /note= "this actively expressed O-fucosyltransferase
FT sequence is claimed for in claim 9"
FT CDS 1. .1098
FT /*tag= a
FT /product= "human heart O-fucosyltransferase"
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.

```
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
DR   P-PSDB; AAW80571.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Claim 9; Fig 12A; 90pp; English.
XX
CC   This DNA encodes a human heart O-fucosyltransferase that can glycosylate
CC   an epidermal growth factor (EGF) domain of a polypeptide with an
CC   activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.
CC   mutants with increased affinity for the EFG domains, are used in
CC   diagnosis and treatment of conditions associated with overexpression of O
CC   -fucosyltransferase, to promote survival of sensory (retinal) neurons.
CC   Probes based on EGF domain polypeptide are used to detect gene
CC   amplification and expression. The expression can also be determined at
CC   the protein level using antibodies specific for O-fucosyltransferase
XX
SQ   Sequence 1514 BP; 320 A; 435 C; 411 G; 348 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                3.08e-200      Length:          1514
Score:                    1979.00        Matches:         365
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       2              Gaps:            0

US-09-774-954-2 (1-365) x AAV65632 (1-1514)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1081 CTGCGGGACGAGTTC 1095

RESULT 2
ACN40746
ID   ACN40746 standard; cDNA; 5218 BP.
XX
AC   ACN40746;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Tumour-associated antigenic target (TAT) cDNA DNA326813, SEQ ID NO:5725.
XX
KW   Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW   tumour; diagnosis; cell proliferative disorder; breast cancer;
KW   colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW   central nervous system cancer; bladder cancer; pancreatic cancer;
KW   cervical cancer; melanoma; leukaemia; hybridisation probe;
KW   chromosome identification; chromosome mapping; gene mapping;
KW   gene therapy; cytostatic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2004030615-A2.
XX
PD   15-APR-2004.
XX
PF   29-SEP-2003; 2003WO-US028547.
XX
PR   02-OCT-2002; 2002US-0414971P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wu TD,  Zhang Z,  Zhou Y;
```

```
XX
DR   WPI; 2004-347921/32.
DR   P-PSDB; ABM82223.
XX
PT   New tumor-associated antigenic target polypeptides and nucleic acids,
PT   useful in preparing a medicament for treating or detecting a
PT   proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT   prostate cancer or tumor.
XX
PS   Claim 1; SEQ ID NO 5725; 7273pp; English.
XX
CC   The invention relates to human tumour-associated antigenic target (TAT)
CC   polypeptides, and their related nucleic acids. The TAT polypeptides are
CC   overexpressed in cancer tissues compared to normal tissues, and may thus
CC   serve as effective targets for the diagnosis and treatment of cancer in
CC   mammals. The invention also relates to nucleic acid and polypeptide
CC   sequences at least 80% identical to the TAT nucleic acids and
CC   polypeptides; expression vectors and host cells comprising a TAT nucleic
CC   acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC   molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC   TAT polypeptide; and methods and compositions for the treatment or
CC   diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC   antibodies, antagonists, binding molecules and compositions are useful
CC   for diagnosing or treating a cell proliferative disorder associated with
CC   increased TAT expression, particularly cancers such as breast cancer,
CC   colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC   cancer, pancreatic cancer, cervical cancer, cancers of the central
CC   nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC   used as hybridisation probes, in chromosome and gene mapping, in
CC   chromosome identification and in gene therapy. The present sequence
CC   represents a TAT nucleic acid of the invention
XX
SQ   Sequence 5218 BP; 1255 A; 1274 C; 1334 G; 1355 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                1.89e-199       Length:          5218
Score:                    1979.00         Matches:         365
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       13              Gaps:            0

US-09-774-954-2 (1-365) x ACN40746 (1-5218)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 598

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 658

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 718

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 778

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 838

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 898

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 958

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1018

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1078

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1138

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1198

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1199 CTGCGGGACGAGTTC 1213

RESULT 3
ADF81754
ID   ADF81754 standard; DNA; 5266 BP.
XX
AC   ADF81754;
XX
DT   26-FEB-2004  (first entry)
XX
DE   Leukaemia-related DNA sequence #2310.
XX
KW   Cytostatic; Gene therapy; leukaemia; ss.
XX
OS   Unidentified.
XX
PN   WO2003039443-A2.
XX
PD   15-MAY-2003.
XX
PF   04-NOV-2002; 2002WO-EP012303.
XX
PR   05-NOV-2001; 2001EP-00126244.
PR   30-APR-2002; 2002EP-00009758.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA   (UYLU-) UNIV LUDWIG MAXIMILIANS.
PA   (HAFE/) HAFERLACH T.
```

PA (SCHO/) SCHOCH C.
PA (KERN/) KERN W.
XX
PI Haferlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;
PI Eils R, Brors B, Mergenthaler S;
XX
DR WPI; 2003-505037/47.
XX
PT Determining the subtype of leukemia cells and whether a patient sample
PT contains leukemia cells or other cells, useful for treating leukemia,
PT comprises determining the expression profile of a group of markers in a
PT patient sample.
XX
PS Disclosure; SEQ ID NO 2310; 2938pp; English.
XX
CC The present invention relates to a method (M1) for determining the
CC subtype of leukaemia cells and whether a patient sample contains
CC leukaemia cells. The method comprises determining the expression profile
CC of a group of markers in a patient sample. The method is useful for
CC determining the presence of leukaemia cells, its types or subtypes, and
CC for the preparation of a medicament for treating leukaemia.
XX
SQ Sequence 5266 BP; 1280 A; 1278 C; 1331 G; 1346 T; 0 U; 31 Other;

Alignment Scores:

Pred. No.:	1.91e-199	Length:	5266
Score:	1979.00	Matches:	365
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

US-09-774-954-2 (1-365) x ADF81754 (1-5266)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 191

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 251

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 311

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        312 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 371

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        372 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 431

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        432 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 491

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        492 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 551

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        552 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 611

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        612 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 671

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        672 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 731

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        732 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 791

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        792 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 851

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        852 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 911

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        912 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 971

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        972 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1031

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1032 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1091

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1092 CTCGGNCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1151

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1152 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1211

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1212 CTGCGGGACGAGTTC 1226
```

RESULT 4
AAV65633
ID AAV65633 standard; DNA; 11284 BP.
XX
AC AAV65633;
XX
DT 16-DEC-1998 (first entry)
XX
DE Plasmid construct for expression of human O-fucosyltransferase.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX

FH	Key	Location/Qualifiers
FT	CDS	4140. .5333
FT		/*tag= a
FT		/note= "insert coding for human O-fucosyltransferase."
FT	misc_feature	4218. .4235
FT		/*tag= b
FT		/note= "polyhistidine tag"

XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;

```
XX
DR   WPI; 1998-437477/37.
DR   P-PSDB; AAW80573.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example; Page 49-57; 90pp; English.
XX
CC   This represents the nucleotide sequence of the plasmid construct used for
CC   the expression of human heart O-fucosyltransferase. The human O-
CC   fucosyltransferase can glycosylate an epidermal growth factor (EGF)
CC   domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC   O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 11284 BP; 2947 A; 2602 C; 2624 G; 3111 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                5.85e-199      Length:          11284
Score:                    1979.00        Matches:         365
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       2              Gaps:            0

US-09-774-954-2 (1-365) x AAV65633 (1-11284)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4416 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 4475

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4476 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 4535

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4536 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 4595

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4596 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 4655

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4656 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 4715

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4716 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 4775

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4776 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 4835

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4836 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 4895

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4896 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 4955

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4956 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 5015

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5016 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 5075

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5076 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 5135

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5136 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 5195

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5196 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 5255

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5256 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 5315

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       5316 CTGCGGGACGAGTTC 5330

RESULT 5
AAK51510
ID   AAK51510 standard; cDNA; 5230 BP.
XX
AC   AAK51510;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 55.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y;
PI   Ma Y,  Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
```

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM78377.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 1; Page 638-642; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 5230 BP; 1285 A; 1255 C; 1330 G; 1360 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.4e-187	Length:	5230
Score:	1867.50	Matches:	350
Percent Similarity:	96.98%	Conservative:	3
Best Local Similarity:	96.15%	Mismatches:	6
Query Match:	94.37%	Indels:	5
DB:	4	Gaps:	1

US-09-774-954-2 (1-365) x AAK51510 (1-5230)

```
Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet----------GlyArgPhe 21
              |||   |||  :::      :::  |||   :::              |||||||||
Db        160 GACATGGCCACGTGGCCTACACATGCTCCAAGCGTGGAGGAAGGAGGCCAGGGGCGCTTT 219

Qy         22 GlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThr 41
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 GGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACC 279

Qy         42 LeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 TTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 339

Qy         62 ValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSer 81
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 GTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGC 399

Qy         82 LeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgValAla 101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 TTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCA 459

Qy        102 TyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGlu 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 TACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAA 519

Qy        122 GlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeu 141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTT 579

Qy        142 PheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerPro 161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 TTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCA 639

Qy        162 LysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnProProValLeuGluGlu 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAA 699

Qy        182 HisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAla 201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 CACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCC 759

Qy        202 GlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAsp 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 CAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGAC 819

Qy        222 TrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSer 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCT 879

Qy        242 ProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeu 261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 CCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTG 939

Qy        262 ProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGln 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 CCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAG 999

Qy        282 SerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLys 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1000 TCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAA 1059

Qy        302 GlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeu 321
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 GGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTC 1119

Qy        322 GlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArg 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1120 GGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGG 1179

Qy        342 GluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLysLeu 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1180 GAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTG 1239

Qy        362 ArgAspGluPhe 365
              |||||||||||
Db       1240 CGGGACGAGTTC 1251
```

RESULT 6
AAK52494
ID AAK52494 standard; cDNA; 4850 BP.
XX
AC AAK52494;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2023.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX

```
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y;
PI   Ma Y,  Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM79361.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 1; Page 4424-4425; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 4850 BP; 1142 A; 1218 C; 1244 G; 1246 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.38e-186      Length:          4850
Score:                  1855.50        Matches:         348
Percent Similarity:     96.70%         Conservative:    4
Best Local Similarity:  95.60%         Mismatches:      7
Query Match:            93.76%         Indels:          5
DB:                     4              Gaps:            1

US-09-774-954-2 (1-365) x AAK52494 (1-4850)

Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet--------------GlyArgPhe 21
              |||   |||   :::     :::   |||   :::              |||||||||
Db        160 GACATGGCCACGTGGCCTACACATGCTCCAAGCGTGGAGGAAGGAGGCCAGGGGCGCTTT 219

Qy         22 GlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThr 41
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        220 GGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTGAACCGTTCC 279

Qy         42 LeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        280 TTGGCTGTCCCTTCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 339

Qy         62 ValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSer 81
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 GTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGC 399

Qy         82 LeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgValAla 101
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 TTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCA 459

Qy        102 TyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGlu 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        460 TACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAA 519

Qy        122 GlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeu 141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        520 GGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTT 579

Qy        142 PheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerPro 161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        580 TTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCA 639

Qy        162 LysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGlu 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        640 AAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAA 699

Qy        182 HisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAla 201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        700 CACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCC 759

Qy        202 GlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAsp 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        760 CAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGAC 819

Qy        222 TrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSer 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        820 TGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCT 879

Qy        242 ProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeu 261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        880 CCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTG 939

Qy        262 ProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGln 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        940 CCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAG 999

Qy        282 SerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLys 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1000 TCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAA 1059

Qy        302 GlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeu 321
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1060 GGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTC 1119

Qy        322 GlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArg 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1120 GGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGG 1179

Qy        342 GluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLysLeu 361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1180 GAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTG 1239

Qy        362 ArgAspGluPhe 365
              |||||||||||
Db       1240 CGGGACGAGTTC 1251

RESULT 7
AAV65634
ID   AAV65634 standard; cDNA; 5009 BP.
XX
AC   AAV65634;
XX
DT   16-DEC-1998  (first entry)
XX
DE   First EcoR1 nucleotide fragment of human KIAA0180.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS   Homo sapiens.
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
DR   P-PSDB; AAW80577.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
```

```
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example 1; Fig 11; 90pp; English.
XX
CC   This represents a first EcoR1 nucleotide fragment of human KIAA0180. This
CC   5009 basepairs partial cDNA encodes for a protein of unknown function
CC   from myeblast cel line KG-1. The invention provides a human heart O-
CC   fucosyltransferase that can glycosylate an epidermal growth factor (EGF)
CC   domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC   O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 5009 BP; 1235 A; 1195 C; 1268 G; 1311 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              4.18e-185       Length:         5009
Score:                  1844.00         Matches:        343
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            93.18%          Indels:         0
DB:                     2               Gaps:           0

US-09-774-954-2 (1-365) x AAV65634 (1-5009)

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 61

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisVal 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGTG 121

Qy         63 SerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeu 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTTG 181

Qy         83 GluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyr 102
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATAC 241

Qy        103 CysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGly 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGGA 301

Qy        123 AsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPhe 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 AACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTTT 361

Qy        143 ThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerProLys 162
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 ACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAAG 421

Qy        163 GluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGluHis 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 GAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACAC 481

Qy        183 ArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAlaGln 202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCAG 541

Qy        203 IleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTrp 222
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 ATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTGG 601

Qy        223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 AAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCG 661

Qy        243 GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro 262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 CAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCT 721

Qy        263 AspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSer 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 GACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCG 781

Qy        283 ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLysGly 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 GTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGG 841

Qy        303 LysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGly 322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 AAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGC 901

Qy        323 GlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArgGlu 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 CAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAG 961

Qy        343 ArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLysLeuArg 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 CGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGG 1021

Qy        363 AspGluPhe 365
              |||||||||
Db       1022 GACGAGTTC 1030

RESULT 8
ABL13947
ID   ABL13947 standard; cDNA; 1209 BP.
XX
AC   ABL13947;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36323.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69844.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 36323; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
```

```
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1209 BP; 302 A; 313 C; 307 G; 287 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                3.58e-78      Length:          1209
Score:                    831.00        Matches:         167
Percent Similarity:       61.23%        Conservative:    62
Best Local Similarity:    44.65%        Mismatches:      119
Query Match:              41.99%        Indels:          26
DB:                       4             Gaps:            9

US-09-774-954-2 (1-365) x ABL13947 (1-1209)

Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGlnAlaAsp 26
              ||||||   |||||||||   ||||||||||||||||||||||||||||||||||||||
Db         82 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCGAC 141

Qy         27 HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProPro 46
              ||||||||||||||||||||||||||||||   ||||||||||||||   :::||||||
Db        142 CACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCG 201

Qy         47 TrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisValSerTyrGlnLys 66
              |||:::||||||:::      :::           :::       |||   :::
Db        202 TGGGTGGAGTATCGT------AGGGGTGAACTGCGATCCCGTCAGGTACCGTTCAACACA 255

Qy         67 TyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeuGluAspPheMet 86
              ||||||:::::|||||||||:::   ||||||||||||||:::::   ||||||||
Db        256 TATTTTGAAGTGGAGCCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTCATG 315

Qy         87 GluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyrCysPheGlu--- 105
                 ||||||          ||||||      :::||||||::::::|||::::::
Db        316 TGGCACCTGGCCGACGACATTTGGCCAGAATCGGAGCGAGTGTCATTTTGCTACAAGGAA 375

Qy        106 ------------ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGlu 121
                         :::      |||||||||      |||      |||:::
Db        376 CGATATAGCCTTCAGCAGGAGAAGAACGATCCAGACAAGCCCAATTGCCACGCCAAGGAT 435

Qy        122 GlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeu 141
              ||||||||||||||||||||||||||   ||||||:::   |||   :::||||||
Db        436 GGCAATCCTTTTGGTCCCTTTTGGGACACTTTTCACATAGACTTTGTGCGGTCAGAGTTC 495

Qy        142 PheThrGlyIleSerPheSerAlaSerTyrArgGluGln---------TrpSerGlnArg 158
              :::      :::   |||          :::      :::          |||      :::
Db        496 TATGCGCCACTTCATTTTGATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCAAA 555

Qy        159 PheSerProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProVal 178
              :::      :::   :::||||||||||||      ||||||||||   |||||||||
Db        556 TATCCTGCAGAATCATATCCCGTACTCGCGTTCACCGGAGCTCCGGCTAGTTTCCTGTT 615

Qy        179 LeuGluGluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThr 198
                 |||:::      ||||||:::|||:::   ||||||          :::
Db        616 CAGCTAGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGTATAGGGAAGCA 675

Qy        199 GlyGluAlaGlnIleHisAlaHisLeuValArg---ProTyrValGlyIleHisLeuArg 217
                 :::      |||          |||   |||      ::::::|||||||||||||||
Db        676 TCTAAGGATTTCATCCGAGAGCAGTTGCCTCGGGGTGCCTTTTTGGGCATTCATCTGCGC 735

Qy        218 IleGlySerAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHis 237
                 |||   ||||||      ||||||      :::||||||                 |||
Db        736 AACGGTATCGATTGGGTGAGAGCCTGTGAGCACGTCAAGGAT---------AGCCAGCAT 786

Qy        238 PheMetAlaSerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMet 257
                 ||||||||||||||:::||||||                 |||   |||
Db        787 CTGTTTGCCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTGCA---CTCTACCCG 843

Qy        258 ThrMetCysLeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSer 277
                 :::|||:::|||      :::   |||   |||   :::|||      :::::::::
Db        844 GAGCTCTGCATGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGAACCATTAAGAAC 903

Qy        278 Leu---------------------AspAlaGlnSerValTyrValAlaThrAspSerGlu 290
                 :::                      :::   :::||||||:::||||||:::||||||
Db        904 GTGCGCCAAACTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCGTCAGACTCCAAT 963

Qy        291 SerTyrValProGluLeuGln---GlnLeuPheLysGlyLysValLysValValSerLeu 309
                 :::   ||||||          |||   :::       :::   |||       |||
Db        964 CACATGATTGGTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGTGTGCACAAGCTG 1023

Qy        310 LysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAspHisPheIleGly 329
                 :::         :::||||||   ||||||||||||:::::||||||||||||
Db       1024 CCGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACCACTTTATCGGC 1083

Qy        330 AsnCysValSerSerPheThrAlaPheValLysArgGluArgAspLeuGlnGlyArgPro 349
              ||||||:::||||||::::::||||||   ||||||||||||||:::   |||   |||
Db       1084 AACTGTATATCCTCTTACTCGGCATTCGAAAAAAGGGAACGAGATGTGCACGGTTTTCCA 1143

Qy        350 SerSerPhePheGlyMetAspArgProProLysLeuArgAsp 363
                 |||   |||:::|||            ||||||   :::|||
Db       1144 TCGTACTTCTGGGGATTC---------CCCAAGGAAAAGGAT 1176

RESULT 9
ADQ56824
ID   ADQ56824 standard; DNA; 610 BP.
XX
AC   ADQ56824;
XX
DT   21-OCT-2004  (first entry)
XX
DE   Novel canine microarray-related DNA sequence SeqID8126.
XX
KW   canine microarray; drug screening; toxicity assay;
KW   environmental pollutant; cellular response; gene expression profile;
KW   toxic response; liver necrosis; fatty liver disease;
KW   protein adduct formation; hepatitis; dog; ds.
XX
OS   Canis familiaris.
XX
PN   WO2004063324-A2.
XX
PD   29-JUL-2004.
XX
PF   05-MAY-2003; 2003WO-US013853.
XX
PR   03-MAY-2002; 2002US-0377240P.
XX
PA   (GENE-) GENE LOGIC INC.
PA   (PFIZ ) PFIZER PROD INC.
XX
PI   Diggans JC,  Porter M,  Wei T;
XX
DR   WPI; 2004-561890/54.
XX
PT   New isolated nucleic acid molecule, useful for drug screening and
PT   toxicity assays or for assessing the impact, including toxicity, of a
PT   compound, pharmaceutical agent or environmental pollutant on a cell or
PT   living organism.
XX
PS   Claim 1; SEQ ID NO 8126; 41pp; English.
XX
CC   This invention is related to a novel isolated canine nucleic acid
CC   sequences and the construction of canine microarrays containing a
CC   significant portion of the canine genome. The isolated canine nucleic
CC   acid sequences of the invention may be useful for drug screening and
CC   toxicity assays. The invention is therefore useful for assessing the
CC   impact, including toxicity, of a compound, pharmaceutical agent or
CC   environmental pollutant on a cell or living organism. The methods are
CC   useful for detecting genes that are up- or down-regulated in canines in a
CC   disease state. The sequences are useful as diagnostic agents or markers
CC   to detect a cellular response in a sample individually or as part of a
CC   gene expression profile. It is also useful as a target for agents that
CC   modulate gene expression or activity. The database is useful for
CC   producing electronic Northerns that allow the user to determine the cell
CC   type or tissue in which a given gene is expressed and to allow
CC   determination of the abundance or expression level of a given gene in a
```

```
CC   particular tissue or cell. The methods are useful for determining the
CC   similarity of a toxic response to one or more individual compounds. The
CC   methods are useful for predicting at least one toxic response or the
CC   likelihood that a compound or test agent will induce various specific
CC   pathologies such as those of the liver (liver necrosis, fatty liver
CC   disease, protein adduct formation or hepatitis), those of the kidney,
CC   heart, brain or testes, or other pathologies associated with at least one
CC   of the toxins. The methods are also useful for predicting or elucidating
CC   the potential cellular pathways influenced, induced or modulated by the
CC   compound or test agent due to the similarity of the expression profile
CC   compared to the profile induced by a known toxin. The present sequence is
CC   that of a canine DNA sequence which was claimed for use during the
CC   production of a canine microarray of the invention.
XX
SQ   Sequence 610 BP; 155 A; 171 C; 149 G; 129 T; 0 U; 6 Other;

Alignment Scores:
Pred. No.:              1.77e-76        Length:         610
Score:                  811.00          Matches:        148
Percent Similarity:     94.55%          Conservative:   8
Best Local Similarity:  89.70%          Mismatches:     9
Query Match:            40.98%          Indels:         1
DB:                     13              Gaps:           0

US-09-774-954-2 (1-365) x ADQ56824 (1-610)

Qy         58 ThrAsnLeuHisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHis 77
              ||||||||||||||    |||||||||||||||||||    ||||||       ||||||
Db          2 ACCAATCTCCATGTGTNCTACCAGAAGTACTTCAAGCTNNAGCCCCTCNAGGNNTACCAT 61

Qy         78 ArgValIleSerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGlu 97
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CGGGTTATCAGCCTAGAGGACTTCATGGAGAAGCTGGCACCCACTCACTGGCCCCCCGAG 121

Qy         98 LysArgValAlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCys 117
              :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 AGGCGAGTGGCATACTGCTTTGAAGTGGCAGCCCAGCGAAGCCCTGATAAGAAGACATGC 181

Qy        118 ProMetLysGluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsn 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 CCCATGAAGGAAGGAAATCCCTTTGGCCCATTTTGGGATCAGTTTCATGTGAGTTTCAAC 241

Qy        138 LysSerGluLeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGln 157
              ||||||:::||||||:::|||||||||||||||||||||||::::::||||||   |||
Db        242 AAGTCCAAGCTTTTTTCAGGCATTTCCTTCAGTGCCTCCTACAAAGACCAATGGATTCAG 301

Qy        158 ArgPheSerProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPhePro 177
              ||||||||||||||||||||||||||||||||    ||||||||||||||||||||||
Db        302 AGATTTTCTCCAAAGGAACATCCAGTGCTGGCCCC-CCGGGAGCCCCCGCCCAGTTCCCC 360

Qy        178 ValLeuGluGluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLys 197
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||:::
Db        361 GTCCTGGAGGAGCACAGACCACTCCAGAAGTATGTGGTATGGTCAGATGAGATGGTGAGG 420

Qy        198 ThrGlyGluAlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArg 217
              ||||||||||||   |||      ||||||:::|||||||||||||||||||||||||||
Db        421 ACGGGAGAGGCCCACATTCTCACCCACCTCATCCGGCCGTACGTGGGCATTCACCTGCGC 480

Qy        218 IleGlySerAspTrp 222
              |||||||||||||||
Db        481 ATTGGCTCTGACTGG 495

RESULT 10
AAD07629
ID   AAD07629 standard; cDNA; 915 BP.
XX
AC   AAD07629;
XX
DT   10-AUG-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 18 cDNA clone HMWEY26, SEQ ID NO:69.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW   endocrine disorder; infection; wound healing; vulnerary; cell culture;
KW   chemotaxis; food additive; binding partner identification; chromosome 3;
KW   ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1. .498
FT                   /*tag=  a
FT                   /product= "Human secreted protein precursor"
FT                   /transl_except= (pos:4. .6, aa:Xaa)
FT                   /transl_except= (pos:10. .12, aa:Xaa)
FT                   /transl_except= (pos:13. .15, aa:Xaa)
FT                   /transl_except= (pos:442. .444, aa:Xaa)
FT                   /note= "Xaa equals any of the naturally occurring L-amino
FT                   acids; CDS does not include start codon"
FT                   /partial
FT   sig_peptide     1. .3
FT                   /*tag=  b
FT   mat_peptide     4. .495
FT                   /*tag=  c
FT                   /product= "Mature human secreted protein"
XX
PN   WO200132676-A1.
XX
PD   10-MAY-2001.
XX
PF   25-OCT-2000; 2000WO-US029365.
XX
PR   29-OCT-1999;   99US-0162237P.
PR   21-JUL-2000; 2000US-0219666P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Shi Y,  Olsen HS,  Soppet DR;
XX
DR   WPI; 2001-328773/34.
DR   P-PSDB; AAE03110.
XX
PT   Nucleic acids encoding 25 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. Gaucher's disease,
PT   Alzheimer's disease, Scimitar syndrome, Creutzfeldt-Jacob disease,
PT   diabetes mellitus and multiple sclerosis.
XX
PS   Claim 1; Page 445; 546pp; English.
XX
CC   AAD07571-AAD07645 represent cDNAs corresponding to 25 human secreted
CC   protein genes, and AAE03052-AAE03126 represent the proteins they encode.
CC   AAE03127-AAE03150 represent human secreted protein fragments. The genes
CC   and their corresponding secreted proteins are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   25 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's
CC   disease), cognitive disorders, schizophrenia, asthma, skin disorders
CC   (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular
CC   disorders, angiogenic disorders, kidney disorders, gastrointestinal
CC   disorders, pregnancy-related disorders, endocrine disorders, and
CC   infections. The proteins can also be used to aid wound healing and
CC   epithelial cell proliferation, to prevent skin aging due to sunburn, to
```

```
CC   maintain organs before transplantation, for supporting cell culture of
CC   primary tissues, to regenerate tissues, to identify their cognate ligands
CC   or binding partners, and in chemotaxis, and can be used as a food
CC   additive or preservative to modify storage properties. Antibodies
CC   specific for a protein of the invention can be used in alleviating
CC   symptoms associated with the disorders mentioned above, and in diagnostic
CC   immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay
CC   (ELISA). The present sequence represents a human secreted protein-
CC   encoding cDNA of the invention
XX
SQ   Sequence 915 BP; 209 A; 253 C; 247 G; 202 T; 0 U; 4 Other;

Alignment Scores:
Pred. No.:              5.23e-76      Length:          915
Score:                  809.00        Matches:         155
Percent Similarity:     96.88%        Conservative:    0
Best Local Similarity:  96.88%        Mismatches:      5
Query Match:            40.88%        Indels:          0
DB:                     4             Gaps:            0

US-09-774-954-2 (1-365) x AAD07629 (1-915)

Qy        206 HisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTrpLysAsnAla 225
              |||     |||   |||   ||||||||||||||||||||||||||||||||||||||||
Db         16 CACTTTGCCCGAACCTACCCGGGCATTCATCTGCGCATTGGCTCTGACTGGAAGAACGCC 75

Qy        226 CysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerProGlnCysVal 245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 TGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCGCAGTGTGTG 135

Qy        246 GlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuProAspLeuLys 265
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 GGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCTGACCTGAAG 195

Qy        266 GluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSerValTyrVal 285
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 GAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCGGTCTACGTT 255

Qy        286 AlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLysGlyLysValLys 305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 GCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAG 315

Qy        306 ValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAsp 325
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 GTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGAC 375

Qy        326 HisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArgGluArgAspLeu 345
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 CACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTC 435

Qy        346 GlnGlyArgProSerSerPhePheGlyMetAspArgProProLysLeuArgAspGluPhe 365
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 CAGGGGAGSCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTC 495

RESULT 11
ABL13946/c
ID   ABL13946 standard; cDNA; 3264 BP.
XX
AC   ABL13946;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36320.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69843.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 36320; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3264 BP; 904 A; 742 C; 758 G; 860 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              8.16e-74      Length:          3264
Score:                  796.00        Matches:         167
Percent Similarity:     58.27%        Conservative:    62
Best Local Similarity:  42.49%        Mismatches:      119
Query Match:            40.22%        Indels:          45
DB:                     4             Gaps:            10

US-09-774-954-2 (1-365) x ABL13946 (1-3264)

Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet---------------------- 18
              ||||||   |||||||||   ||||||||||||||
Db       2183 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGTAAGTTTCACTTGGAAGCAACC 2124

Qy         19 ------------------------------GlyArgPheGlyAsnGlnAlaAspHisPh 28
                                            ||||||||||||||||||||||||||||||
Db       2123 GAAAAACGACAACTAAACCGCGCTTTTCAAAGGACGCTTTGGCAACCAGGCCGACCACTT 2064

Qy         28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProProTrpIl 48
              ||||||||||||||||||||||||   ||||||||||||||   :::||||||||::
Db       2063 CCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCGTGGGT 2004

Qy         48 eGluTyrGlnHisHisLysProProPheThrAsnLeuHisValSerTyrGlnLysTyrPh 68
              :|||||:::      :::          :::          |||   :::      |||||
Db       2003 GGAGTATCGT------AGGGGTGAACTGCGATCCCGTCAGGTACCGTTCAACACATATTT 1950

Qy         68 eLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeuGluAspPheMetGluLy 88
              |::::::||||||||:::   ||||||||||||||:::::   |||||||||
Db       1949 TGAAGTGGAGCCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTCATGTGGCA 1890

Qy         88 sLeuAlaProThrHisTrpProProGluLysArgValAlaTyrCysPheGlu-------- 105
              ||||||         ||||||      :::||||||::::::|||::::::
Db       1889 CCTGGCCGACGACATTTGGCCAGAATCGGAGCGAGTGTCATTTTGCTACAAGGAACGATA 1830

Qy        106 -------ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGlyAs 123
                     :::      |||||||||      |||      |||:::|||||
Db       1829 TAGCCTTCAGCAGGAGAAGAACGATCCAGACAAGCCCAATTGCCACGCCAAGGATGGCAA 1770

Qy        123 nProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPheTh 143
              ||||||||||||||||||||   ||||||:::   |||   :::|||||||   :::
Db       1769 TCCTTTTGGTCCCTTTTGGGACACTTTTCACATAGACTTTGTGCGGTCAGAGTTCTATGC 1710
```

```
Qy        143 rGlyIleSerPheSerAlaSerTyrArgGluGln---------TrpSerGlnArgPheSe 160
Db       1709 GCCACTTCATTTTGATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCAAATATCC 1650

Qy        160 rProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGl 180
Db       1649 TGCAGAATCATATCCCGTACTCGCGTTCACCGGAGCTCCGGCTAGTTTTCCTGTTCAGCT 1590

Qy        180 uGluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGl 200
Db       1589 AGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGTATAGGGAAGCATCTAA 1530

Qy        200 uAlaGlnIleHisAlaHisLeuValArg---ProTyrValGlyIleHisLeuArgIleGl 219
Db       1529 GGATTTCATCCGAGAGCAGTTGCCTCGGGGTGCCTTTTTGGGCATTCATCTGCGCAACGG 1470

Qy        219 ySerAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMe 239
Db       1469 TATCGATTGGGTGAGAGCCTGTGAGCACGTCAAGGAT---------AGCCAGCATCTGTT 1419

Qy        239 tAlaSerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMe 259
Db       1418 TGCCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTGCA---CTCTACCCGGAGCT 1362

Qy        259 tCysLeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeu-- 278
Db       1361 CTGCATGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGAACCATTAAGAACGTGCG 1302

Qy        279 ------------------AspAlaGlnSerValTyrValAlaThrAspSerGluSerTy 292
Db       1301 CCAAACTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCGTCAGACTCCAATCACAT 1242

Qy        292 rValProGluLeuGln---GlnLeuPheLysGlyLysValLysValSerLeuLysPr 311
Db       1241 GATTGGTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGTGTGCACAAGCTGCCGGA 1182

Qy        311 oGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAspHisPheIleGlyAsnCy 331
Db       1181 GGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACCACTTTATCGGCAACTG 1122

Qy        331 sValSerSerPheThrAlaPheValLysArgGluArgAspLeuGlnGlyArgProSerSe 351
Db       1121 TATATCCTCTTACTCGGCATTCGAAAAAAGGGAACGAGATGTGCACGGTTTTCCATCGTA 1062

Qy        351 rPhePheGlyMetAspArgProProLysLeuArgAsp 363
Db       1061 CTTCTGGGGATTC---------CCCAAGGAAAAGGAT 1034
```

RESULT 12
ABL13894/c
ID ABL13894 standard; cDNA; 3793 BP.
XX
AC ABL13894;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 36164.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.

XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB69791.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Claim 1; SEQ ID NO 36164; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 3793 BP; 1020 A; 853 C; 950 G; 970 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	6.69e-43	Length:	3793
Score:	506.50	Matches:	103
Percent Similarity:	56.38%	Conservative:	34
Best Local Similarity:	42.39%	Mismatches:	59
Query Match:	25.59%	Indels:	47
DB:	4	Gaps:	5

US-09-774-954-2 (1-365) x ABL13894 (1-3793)

```
Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet----------------------- 18
Db        675 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGTAAGTTTCACTTGGAAGCAACC 616

Qy         19 ------------------------------GlyArgPheGlyAsnGlnAlaAspHisPh 28
Db        615 GAAAAACGACAACTAAACCGCGCTTTTCAAAGGACGCTTTGGCAACCAGGCCGACCACTT 556

Qy         28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProProTrpIl 48
Db        555 CCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCGTGGGT 496

Qy         48 eGluTyrGlnHisHisLysProProPheThrAsnLeuHisValSerTyrGlnLysTyrPh 68
Db        495 GGAGTATCGT------AGGGGTGAACTGCGATCCCGTCAGGTACCGTTCAACACATATTT 442

Qy         68 eLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeuGluAspPheMetGluLy 88
Db        441 TGAAGTGGAGCCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTCATGTGGCA 382

Qy         88 sLeuAlaProThrHisTrpProProGluLysArgValAlaTyrCysPheGlu-------- 105
Db        381 CCTGGCCGACGACATTTGGCCAGAATCGGAGCGAGTGTCATTTTGCTACAAGGAACGATA 322

Qy        106 -------ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGlyAs 123
Db        321 TAGCCTTCAGCAGGAGAAGAACGATCCAGACAAGCCCAATTGCCACGCCAAGGATGGCAA 262

Qy        123 nProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPheTh 143
Db        261 TCCTTTTGGTCCCTTTTGGGACACTTTTCACATAGACTTTGTGCGGTCAGAGTTCTATGC 202

Qy        143 rGlyIleSerPheSerAlaSerTyrArgGluGln---------TrpSerGlnArgPheSe 160
Db        201 GCCACTTCATTTTGATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCAAATATCC 142

Qy        160 rProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGl 180
```

```
Db        141 TGCAGAATCATATCCCGTACTCGCGTTCACCGGAGCTCCGGCTAGTTTTCCTGTTCAGCT 82

Qy        180 uGluHisArgProLeuGlnLysTyrMetValTrpSer-AspGluMetValLysThrGlyG 200
              |||:::      ||||||:::|||:::   |||||| :::
Db         81 AGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGT---------------- 38

Qy        200 luAlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlyS 220
                                          :::|||   ||||||||||||   |
Db         37 ----------------------------------ATAGGGAAGCATCTAAGGATTTCAT 13

Qy        220 erAsp 221
              ||:::
Db         12 CCGAG 8
```

RESULT 13
AAK79915
ID AAK79915 standard; DNA; 10331 BP.
XX
AC AAK79915;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34727.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.

```
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 34727; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 10331 BP; 2718 A; 2405 C; 2599 G; 2609 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.76e-33        Length:         10331
Score:                  424.00          Matches:        83
Percent Similarity:     98.81%          Conservative:   0
Best Local Similarity:  98.81%          Mismatches:     1
Query Match:            21.42%          Indels:         0
DB:                     4               Gaps:           0

US-09-774-954-2 (1-365) x AAK79915 (1-10331)

Qy        223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2491 AAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCG 2550

Qy        243 GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro 262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       2551 CAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCT 2610

Qy        263 AspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSer 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2611 GACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCG 2670

Qy        283 ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLysGly 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2671 GTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGG 2730

Qy        303 LysValLysVal 306
              ||||||   |||
Db       2731 AAGGTATGTGTG 2742

RESULT 14
AAL03344
ID   AAL03344 standard; DNA; 10331 BP.
XX
AC   AAL03344;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 6032.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001339.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
```

PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; SEQ ID NO 6032; 1297pp + Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 10331 BP; 2718 A; 2405 C; 2599 G; 2609 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.76e-33	Length:	10331
Score:	424.00	Matches:	83
Percent Similarity:	98.81%	Conservative:	0
Best Local Similarity:	98.81%	Mismatches:	1
Query Match:	21.42%	Indels:	0
DB:	4	Gaps:	0

US-09-774-954-2 (1-365) x AAL03344 (1-10331)

```
Qy       223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2491 AAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCG 2550

Qy       243 GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro 262
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2551 CAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCT 2610

Qy       263 AspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSer 282
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2611 GACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCG 2670

Qy       283 ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLysGly 302
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2671 GTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGG 2730

Qy       303 LysValLysVal 306
            ||||||   |||
```

```
Db       2731 AAGGTATGTGTG 2742

RESULT 15
ACH32225
ID   ACH32225 standard; cDNA; 479 BP.
XX
AC   ACH32225;
XX
DT   13-OCT-2003  (first entry)
XX
DE   Human endothelial cell cDNA #358.
XX
KW   Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW   genome mapping; biodiversity; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   US2003073623-A1.
XX
PD   17-APR-2003.
XX
PF   30-JUL-2001; 2001US-00918995.
XX
PR   30-JUL-2001; 2001US-00918995.
XX
PA   (DRMA/) DRMANAC R T.
PA   (LABA/) LABAT I.
PA   (STAC/) STACHE-CRAIN B.
PA   (DICK/) DICKSON M C.
PA   (JONE/) JONES L W.
XX
PI   Drmanac RT,  Labat I,  Stache-Crain B,  Dickson MC,  Jones LW;
XX
DR   WPI; 2003-615964/58.
XX
PT   New polynucleotide sequences obtained from various cDNA libraries, useful
PT   as hybridization probes, as oligomers for PCR, for chromosome and gene
PT   mapping, in the recombinant production of protein, or in generating
PT   antisense DNA or RNA.
XX
PS   Claim 1; SEQ ID NO 19437; 44pp; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any one of
CC   38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC   determined by the technique of SBH (sequencing by hybridisation). Also
CC   included is a purified polypeptide comprising a sequence corresponding to
CC   a reading frame of the novel polynucleotide. The nucleic acid sequences
CC   are useful in diagnostics as expressed sequence tags (EST) for
CC   identifying expressed genes or for physical mapping of the human genome,
CC   in forensics, in assessing biodiversities, or in identifying mutations
CC   responsible for genetic disorders and other traits. The nucleotide
CC   sequences are also useful as hybridisation probes, as oligomers for PCR,
CC   for chromosome and gene mapping, in the recombinant production of
CC   protein, or in generating antisense DNA or RNA. The purified polypeptide
CC   is useful for generating antibodies specific for it. The present sequence
CC   is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ   Sequence 479 BP; 99 A; 140 C; 143 G; 97 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                3.9e-32        Length:          479
Score:                    393.00         Matches:         76
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              19.86%         Indels:          0
DB:                       9              Gaps:            0

US-09-774-954-2 (1-365) x ACH32225 (1-479)

Qy       290 GluSerTyrValProGluLeuGlnGlnLeuPheLysGlyLysValLysValValSerLeu 309
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        31 GAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTG 90

Qy       310 LysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAspHisPheIleGly 329
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        91 AAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGC 150

Qy       330 AsnCysValSerSerPheThrAlaPheValLysArgGluArgAspLeuGlnGlyArgPro 349
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       151 AACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCG 210

Qy       350 SerSerPhePheGlyMetAspArgProProLysLeuArgAspGluPhe 365
                |||||||||||||||||||||||||||||||||||||||||||||
Db       211 TCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTC 258

Search completed: October 26, 2005, 09:51:26
Job time : 1056.3 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 09:31:14 ; Search time 1559.68 Seconds
(without alignments)
1931.299 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....QGRPSSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9772377 seqs, 4126317084 residues

Total number of hits satisfying chosen parameters: 19544754

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105434_6433/app_query.fasta_1.917
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09774954_@CGN_1_1_684_@runat_25102005_105434_6433
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1979	100.0	1100	11	US-09-774-954-16	Sequence 16, Appl
2	1979	100.0	1300	11	US-09-774-954-4	Sequence 4, Appli
3	1979	100.0	1514	11	US-09-774-954-1	Sequence 1, Appli
4	1979	100.0	5218	16	US-10-301-822-168	Sequence 168, App
5	1979	100.0	5218	22	US-10-956-157-4752	Sequence 4752, Ap
6	1979	100.0	11284	11	US-09-774-954-5	Sequence 5, Appli
7	1844	93.2	5009	11	US-09-774-954-7	Sequence 7, Appli
8	889	44.9	1400	22	US-10-956-157-9986	Sequence 9986, Ap
9	889	44.9	1506	22	US-10-956-157-4751	Sequence 4751, Ap
10	831	42.0	1209	26	US-11-097-143-18161	Sequence 18161, A
c 11	796	40.2	3264	26	US-11-097-143-18160	Sequence 18160, A
c 12	506.5	25.6	3793	26	US-11-097-143-18082	Sequence 18082, A
13	424	21.4	10331	10	US-09-764-891-6032	Sequence 6032, Ap
14	393	19.9	479	10	US-09-918-995-19437	Sequence 19437, A
c 15	392	19.8	3567	26	US-11-097-143-18169	Sequence 18169, A
16	353	17.8	477	10	US-09-918-995-22744	Sequence 22744, A
17	264	13.3	402	9	US-09-783-590-11501	Sequence 11501, A
18	164.5	8.3	1967	20	US-10-471-450-52	Sequence 52, Appl
19	164	8.3	1926	26	US-11-097-143-36191	Sequence 36191, A
c 20	164	8.3	3926	26	US-11-097-143-36190	Sequence 36190, A
21	142	7.2	2434	20	US-10-471-450-34	Sequence 34, Appl
c 22	120	6.1	521	13	US-09-925-065A-602444	Sequence 602444,
23	120	6.1	760	24	US-10-820-474A-236	Sequence 236, App
24	117	5.9	2238	19	US-10-424-599-2236	Sequence 2236, Ap
c 25	115.5	5.8	4104	26	US-11-097-143-36187	Sequence 36187, A
26	110	5.6	1574	22	US-10-764-420-2104	Sequence 2104, Ap
27	109	5.5	1551	18	US-10-062-674-1876	Sequence 1876, Ap
28	108	5.5	3052	22	US-10-956-157-4145	Sequence 4145, Ap
29	106.5	5.4	2166	20	US-10-437-963-59638	Sequence 59638, A
30	105.5	5.3	1512	22	US-10-956-157-1986	Sequence 1986, Ap
31	103.5	5.2	960	18	US-10-309-290-119	Sequence 119, App
32	103.5	5.2	960	18	US-10-309-290-121	Sequence 121, App
33	103	5.2	1960	19	US-10-425-114-31610	Sequence 31610, A
34	103	5.2	2122	21	US-10-425-115-143813	Sequence 143813,
35	102.5	5.2	783	16	US-10-128-714-2312	Sequence 2312, Ap
36	102.5	5.2	907	16	US-10-128-714-1312	Sequence 1312, Ap
37	102.5	5.2	2907	16	US-10-128-714-312	Sequence 312, App
c 38	102	5.2	2790	20	US-10-437-963-71433	Sequence 71433, A
39	102	5.2	29069	24	US-10-915-740A-44	Sequence 44, Appl
40	102	5.2	2242716	24	US-10-915-740A-1068	Sequence 1068, Ap
41	101	5.1	1617	20	US-10-437-963-11436	Sequence 11436, A
c 42	101	5.1	2600	24	US-10-450-763-10656	Sequence 10656, A
c 43	100.5	5.1	24678	22	US-10-741-600-17702	Sequence 17702, A
44	99	5.0	14399	19	US-10-287-226-321	Sequence 321, App
45	99	5.0	140152	21	US-10-684-422-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1
US-09-774-954-16
; Sequence 16, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA

```
;           ZIP: 94080
;        COMPUTER READABLE FORM:
;           MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/09/774,954
;           FILING DATE: 30-Jan-2001
;           CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US/08/978,741
;           FILING DATE: 26-NOV-1997
;           APPLICATION NUMBER: 08/792,498
;           FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;           NAME: Svoboda, Craig G.
;           REGISTRATION NUMBER: 39,044
;           REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;           TELEPHONE: 650/225-1489
;           TELEFAX: 650/952-9881
;     INFORMATION FOR SEQ ID NO: 16:
;        SEQUENCE CHARACTERISTICS:
;           LENGTH: 1100 base pairs
;           TYPE: Nucleic Acid
;           STRANDEDNESS: Single
;           TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-774-954-16

Alignment Scores:
Pred. No.:                1.27e-238      Length:          1100
Score:                    1979.00        Matches:         365
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       11             Gaps:            0

US-09-774-954-2 (1-365) x US-09-774-954-16 (1-1100)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1081 CTGCGGGACGAGTTC 1095

RESULT 2
US-09-774-954-4
; Sequence 4, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;           ADDRESSEE: Genentech, Inc.
;           STREET: 1 DNA Way
;           CITY: South San Francisco
;           STATE: California
;           COUNTRY: USA
;           ZIP: 94080
;        COMPUTER READABLE FORM:
;           MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/09/774,954
;           FILING DATE: 30-Jan-2001
;           CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US/08/978,741
;           FILING DATE: 26-NOV-1997
;           APPLICATION NUMBER: 08/792,498
;           FILING DATE: 31-JAN-1997
```

```
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Svoboda, Craig G.
;              REGISTRATION NUMBER: 39,044
;              REFERENCE/DOCKET NUMBER: P1041P1
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 650/225-1489
;              TELEFAX: 650/952-9881
;    INFORMATION FOR SEQ ID NO: 4:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1300 base pairs
;              TYPE: Nucleic Acid
;              STRANDEDNESS: Single
;              TOPOLOGY: Linear
;         SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-774-954-4

Alignment Scores:
Pred. No.:                1.63e-238       Length:          1300
Score:                    1979.00         Matches:         365
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       11              Gaps:            0

US-09-774-954-2 (1-365) x US-09-774-954-4 (1-1300)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1216 CTGCGGGACGAGTTC 1230

RESULT 3
US-09-774-954-1
; Sequence 1, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;         APPLICANT: Yang Wang, Michael W. Spellman
;         TITLE OF INVENTION: O-Fucosyltransferase
;         NUMBER OF SEQUENCES: 17
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: Genentech, Inc.
;              STREET: 1 DNA Way
;              CITY: South San Francisco
;              STATE: California
;              COUNTRY: USA
;              ZIP: 94080
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: WinPatin (Genentech)
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/774,954
;              FILING DATE: 30-Jan-2001
;              CLASSIFICATION: <Unknown>
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: US/08/978,741
;              FILING DATE: 26-NOV-1997
;              APPLICATION NUMBER: 08/792,498
;              FILING DATE: 31-JAN-1997
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Svoboda, Craig G.
;              REGISTRATION NUMBER: 39,044
;              REFERENCE/DOCKET NUMBER: P1041P1
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 650/225-1489
;              TELEFAX: 650/952-9881
;    INFORMATION FOR SEQ ID NO: 1:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 1514 base pairs
;              TYPE: Nucleic Acid
;              STRANDEDNESS: Single
;              TOPOLOGY: Linear
;         SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-774-954-1
```

```
Alignment Scores:
Pred. No.:              2.05e-238       Length:         1514
Score:                  1979.00         Matches:        365
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     11              Gaps:           0

US-09-774-954-2 (1-365) x US-09-774-954-1 (1-1514)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnProProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1081 CTGCGGGACGAGTTC 1095

RESULT 4
US-10-301-822-168
; Sequence 168, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
;  APPLICANT:  Burgart, Lawrence J.
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
;  CURRENT APPLICATION NUMBER: US/10/301,822
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 168
;   LENGTH: 5218
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (50)...(1216)
US-10-301-822-168

Alignment Scores:
Pred. No.:              1.31e-237       Length:         5218
Score:                  1979.00         Matches:        365
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     16              Gaps:           0

US-09-774-954-2 (1-365) x US-10-301-822-168 (1-5218)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy          81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy         101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy         121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy         141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 598

Qy         161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         599 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 658

Qy         181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         659 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 718

Qy         201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         719 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 778

Qy         221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         779 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 838

Qy         241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         839 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 898

Qy         261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         899 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 958

Qy         281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         959 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1018

Qy         301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1019 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1078

Qy         321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1079 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1138

Qy         341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1139 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1198

Qy         361 LeuArgAspGluPhe 365
               |||||||||||||||
Db        1199 CTGCGGGACGAGTTC 1213
```

RESULT 5
US-10-956-157-4752
; Sequence 4752, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4752
; LENGTH: 5218
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-4752

Alignment Scores:

Pred. No.:	1.31e-237	Length:	5218
Score:	1979.00	Matches:	365
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0

US-09-774-954-2 (1-365) x US-10-956-157-4752 (1-5218)

```
Qy           1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy          21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy          41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy          61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy          81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy         101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy         121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy         141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 598

Qy         161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         599 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 658

Qy         181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         659 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 718

Qy         201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         719 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 778

Qy         221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         779 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 838

Qy         241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         839 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 898

Qy         261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       899 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 958

Qy       281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       959 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1018

Qy       301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1019 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1078

Qy       321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1079 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1138

Qy       341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1139 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1198

Qy       361 LeuArgAspGluPhe 365
             |||||||||||||||
Db      1199 CTGCGGGACGAGTTC 1213
```

RESULT 6
US-09-774-954-5
; Sequence 5, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11284 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-774-954-5

Alignment Scores:

Pred. No.:	4.19e-237	Length:	11284
Score:	1979.00	Matches:	365
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	11	Gaps:	0

US-09-774-954-2 (1-365) x US-09-774-954-5 (1-11284)

```
Qy         1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy        21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy        41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy        61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4416 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 4475

Qy        81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4476 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 4535

Qy       101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4536 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 4595

Qy       121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4596 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 4655

Qy       141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4656 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 4715

Qy       161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4716 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 4775

Qy       181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4776 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 4835

Qy       201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4836 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 4895

Qy       221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4896 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 4955

Qy       241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4956 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 5015

Qy       261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5016 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 5075

Qy       281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5076 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 5135

Qy       301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5136 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 5195

Qy       321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      5196 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 5255
```

```
Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5256 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 5315

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       5316 CTGCGGGACGAGTTC 5330
```

```
RESULT 7
US-09-774-954-7
; Sequence 7, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 7:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 5009 base pairs
;             TYPE: Nucleic Acid
;             STRANDEDNESS: Single
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-774-954-7

Alignment Scores:
Pred. No.:              1.32e-220       Length:         5009
Score:                  1844.00         Matches:        343
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            93.18%          Indels:         0
DB:                     11              Gaps:           0

US-09-774-954-2 (1-365) x US-09-774-954-7 (1-5009)

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 61

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisVal 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGTG 121

Qy         63 SerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeu 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTTG 181

Qy         83 GluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyr 102
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATAC 241

Qy        103 CysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGly 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGGA 301

Qy        123 AsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPhe 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 AACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTTT 361

Qy        143 ThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerProLys 162
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 ACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAAG 421

Qy        163 GluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGluHis 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 GAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACAC 481

Qy        183 ArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAlaGln 202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCAG 541

Qy        203 IleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTrp 222
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 ATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTGG 601

Qy        223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 AAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCG 661

Qy        243 GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro 262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 CAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCT 721

Qy        263 AspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSer 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 GACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCG 781

Qy        283 ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLysGly 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 GTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGG 841

Qy        303 LysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGly 322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 AAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGC 901

Qy        323 GlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArgGlu 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 CAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAG 961

Qy        343 ArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLysLeuArg 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 CGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGG 1021

Qy        363 AspGluPhe 365
              |||||||||
Db       1022 GACGAGTTC 1030
```

```
RESULT 8
US-10-956-157-9986
; Sequence 9986, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William
;  TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;  TITLE OF INVENTION:  HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;  FILE REFERENCE: 031896-043000 (AM 101081)
;  CURRENT APPLICATION NUMBER: US/10/956,157
```

```
;  CURRENT FILING DATE:  2004-10-04
;  NUMBER OF SEQ ID NOS: 319805
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9986
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-9986

Alignment Scores:
Pred. No.:              5.67e-101       Length:         1400
Score:                  889.00          Matches:        158
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            44.92%          Indels:         0
DB:                     22              Gaps:           0

US-09-774-954-2 (1-365) x US-10-956-157-9986 (1-1400)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 72

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 132

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 192

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 252

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 312

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 372

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 432

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArg 158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGG 486

RESULT 9
US-10-956-157-4751
; Sequence 4751, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William
;  TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;  TITLE OF INVENTION:  HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;  FILE REFERENCE: 031896-043000 (AM 101081)
;  CURRENT APPLICATION NUMBER: US/10/956,157
;  CURRENT FILING DATE:  2004-10-04
;  NUMBER OF SEQ ID NOS: 319805
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4751
;   LENGTH: 1506
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-4751

Alignment Scores:
Pred. No.:              6.33e-101       Length:         1506
Score:                  889.00          Matches:        158
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            44.92%          Indels:         0
DB:                     22              Gaps:           0

US-09-774-954-2 (1-365) x US-10-956-157-4751 (1-1506)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 358

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 418

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 478

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 538

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArg 158
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGG 592

RESULT 10
US-11-097-143-18161
; Sequence 18161, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18161
;   LENGTH: 1209
```

```
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-18161

Alignment Scores:
Pred. No.:              9.42e-94      Length:        1209
Score:                  831.00        Matches:       167
Percent Similarity:     61.23%        Conservative:  62
Best Local Similarity:  44.65%        Mismatches:    119
Query Match:            41.99%        Indels:        26
DB:                     26            Gaps:          9

US-09-774-954-2 (1-365) x US-11-097-143-18161 (1-1209)

Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGlnAlaAsp 26
Db         82 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCGAC 141

Qy         27 HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProPro 46
Db        142 CACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCG 201

Qy         47 TrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisValSerTyrGlnLys 66
Db        202 TGGGTGGAGTATCGT------AGGGGTGAACTGCGATCCCGTCAGGTACCGTTCAACACA 255

Qy         67 TyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeuGluAspPheMet 86
Db        256 TATTTGAAGTGGAGCCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTCATG 315

Qy         87 GluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyrCysPheGlu--- 105
Db        316 TGGCACCTGGCCGACGACATTTGGCCAGAATCGGAGCGAGTGTCATTTTGCTACAAGGAA 375

Qy        106 -----------ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGlu 121
Db        376 CGATATAGCCTTCAGCAGGAGAAGAACGATCCAGACAAGCCCAATTGCCACGCCAAGGAT 435

Qy        122 GlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeu 141
Db        436 GGCAATCCTTTTGGTCCCTTTTGGGACACTTTTCACATAGACTTTGTGCGGTCAGAGTTC 495

Qy        142 PheThrGlyIleSerPheSerAlaSerTyrArgGluGln---------TrpSerGlnArg 158
Db        496 TATGCGCCACTTCATTTTGATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCAAA 555

Qy        159 PheSerProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProVal 178
Db        556 TATCCTGCAGAATCATATCCCGTACTCGCGTTCACCGGAGCTCCGGCTAGTTTTCCTGTT 615

Qy        179 LeuGluGluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThr 198
Db        616 CAGCTAGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGTATAGGGAAGCA 675

Qy        199 GlyGluAlaGlnIleHisAlaHisLeuValArg---ProTyrValGlyIleHisLeuArg 217
Db        676 TCTAAGGATTTCATCCGAGAGCAGTTGCCTCGGGGTGCCTTTTGGGCATTCATCTGCGC 735

Qy        218 IleGlySerAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHis 237
Db        736 AACGGTATCGATTGGGTGAGAGCCTGTGAGCACGTCAAGGAT---------AGCCAGCAT 786

Qy        238 PheMetAlaSerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMet 257
Db        787 CTGTTTGCCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTGCA---CTCTACCCG 843

Qy        258 ThrMetCysLeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSer 277
Db        844 GAGCTCTGCATGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGAACCATTAAGAAC 903

Qy        278 Leu--------------------AspAlaGlnSerValTyrValAlaThrAspSerGlu 290
Db        904 GTGCGCCAAACTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCGTCAGACTCCAAT 963

Qy        291 SerTyrValProGluLeuGln---GlnLeuPheLysGlyLysValLysValValSerLeu 309
Db        964 CACATGATTGGTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGTGTGCACAAGCTG 1023

Qy        310 LysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAspHisPheIleGly 329
Db       1024 CCGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACCACTTTATCGGC 1083

Qy        330 AsnCysValSerSerPheThrAlaPheValLysArgGluArgAspLeuGlnGlyArgPro 349
Db       1084 AACTGTATATCCTCTTACTCGGCATTCGAAAAAAGGGAACGAGATGTGCACGGTTTTCCA 1143

Qy        350 SerSerPhePheGlyMetAspArgProProLysLeuArgAsp 363
Db       1144 TCGTACTTCTGGGGATTC---------CCCAAGGAAAAGGAT 1176

RESULT 11
US-11-097-143-18160/c
; Sequence 18160, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18160
;   LENGTH: 3264
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-18160

Alignment Scores:
Pred. No.:              1.09e-88      Length:        3264
Score:                  796.00        Matches:       167
Percent Similarity:     58.27%        Conservative:  62
Best Local Similarity:  42.49%        Mismatches:    119
Query Match:            40.22%        Indels:        45
DB:                     26            Gaps:          10

US-09-774-954-2 (1-365) x US-11-097-143-18160 (1-3264)

Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet------------------------ 18
Db       2183 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGTAAGTTTCACTTGGAAGCAACC 2124

Qy         19 ------------------------------GlyArgPheGlyAsnGlnAlaAspHisPh 28
Db       2123 GAAAAACGACAACTAAACCGCGCTTTTCAAAGGACGCTTTGGCAACCAGGCCGACCACTT 2064

Qy         28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProProTrpIl 48
```

```
Db       2063 CCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCGTGGGT 2004

Qy         48 eGluTyrGlnHisHisLysProProPheThrAsnLeuHisValSerTyrGlnLysTyrPh 68

Db       2003 GGAGTATCGT------AGGGGTGAACTGCGATCCCGTCAGGTACCGTTCAACACATATTT 1950

Qy         68 eLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeuGluAspPheMetGluLy 88

Db       1949 TGAAGTGGAGCCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTCATGTGGCA 1890

Qy         88 sLeuAlaProThrHisTrpProProGluLysArgValAlaTyrCysPheGlu-------- 105

Db       1889 CCTGGCCGACGACATTTGGCCAGAATCGGAGCGAGTGTCATTTTGCTACAAGGAACGATA 1830

Qy        106 -------ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGlyAs 123

Db       1829 TAGCCTTCAGCAGGAGAAGAACGATCCAGACAAGCCCAATTGCCACGCCAAGGATGGCAA 1770

Qy        123 nProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPheTh 143

Db       1769 TCCTTTTGGTCCCTTTTGGGACACTTTTCACATAGACTTTGTGCGGTCAGAGTTCTATGC 1710

Qy        143 rGlyIleSerPheSerAlaSerTyrArgGluGln---------TrpSerGlnArgPheSe 160

Db       1709 GCCACTTCATTTTGATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCAAATATCC 1650

Qy        160 rProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGl 180

Db       1649 TGCAGAATCATATCCCGTACTCGCGTTCACCGGAGCTCCGGCTAGTTTTCCTGTTCAGCT 1590

Qy        180 uGluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGl 200

Db       1589 AGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGTATAGGGAAGCATCTAA 1530

Qy        200 uAlaGlnIleHisAlaHisLeuValArg---ProTyrValGlyIleHisLeuArgIleGl 219

Db       1529 GGATTTCATCCGAGAGCAGTTGCCTCGGGGTGCCTTTTTGGGCATTCATCTGCGCAACGG 1470

Qy        219 ySerAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMe 239

Db       1469 TATCGATTGGGTGAGAGCCTGTGAGCACGTCAAGGAT---------AGCCAGCATCTGTT 1419

Qy        239 tAlaSerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMe 259

Db       1418 TGCCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTGCA---CTCTACCCGGAGCT 1362

Qy        259 tCysLeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeu-- 278

Db       1361 CTGCATGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGAACCATTAAGAACGTGCG 1302

Qy        279 ------------------AspAlaGlnSerValTyrValAlaThrAspSerGluSerTy 292

Db       1301 CCAAACTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCGTCAGACTCCAATCACAT 1242

Qy        292 rValProGluLeuGln---GlnLeuPheLysGlyLysValLysValValSerLeuLysPr 311

Db       1241 GATTGGTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGTGTGCACAAGCTGCCGGA 1182

Qy        311 oGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAspHisPheIleGlyAsnCy 331

Db       1181 GGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACCACTTTATCGGCAACTG 1122

Qy        331 sValSerSerPheThrAlaPheValLysArgGluArgAspLeuGlnGlyArgProSerSe 351

Db       1121 TATATCCTCTTACTCGGCATTCGAAAAAAGGGAACGAGATGTGCACGGTTTTCCATCGTA 1062

Qy        351 rPhePheGlyMetAspArgProProLysLeuArgAsp 363

Db       1061 CTTCTGGGGATTC---------CCCAAGGAAAAGGAT 1034

RESULT 12
US-11-097-143-18082/c
```

```
; Sequence 18082, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18082
;   LENGTH: 3793
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-18082

Alignment Scores:
Pred. No.:                4.49e-52        Length:          3793
Score:                    506.50          Matches:         103
Percent Similarity:       56.38%          Conservative:    34
Best Local Similarity:    42.39%          Mismatches:      59
Query Match:              25.59%          Indels:          47
DB:                       26              Gaps:            5

US-09-774-954-2 (1-365) x US-11-097-143-18082 (1-3793)

Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet----------------------- 18

Db        675 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGTAAGTTTCACTTGGAAGCAACC 616

Qy         19 ------------------------------GlyArgPheGlyAsnGlnAlaAspHisPh 28

Db        615 GAAAAACGACAACTAAACCGCGCTTTTCAAAGGACGCTTTGGCAACCAGGCCGACCACTT 556

Qy         28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProProTrpIl 48

Db        555 CCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCGTGGGT 496

Qy         48 eGluTyrGlnHisHisLysProProPheThrAsnLeuHisValSerTyrGlnLysTyrPh 68

Db        495 GGAGTATCGT------AGGGGTGAACTGCGATCCCGTCAGGTACCGTTCAACACATATTT 442

Qy         68 eLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeuGluAspPheMetGluLy 88

Db        441 TGAAGTGGAGCCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTCATGTGGCA 382

Qy         88 sLeuAlaProThrHisTrpProProGluLysArgValAlaTyrCysPheGlu-------- 105

Db        381 CCTGGCCGACGACATTTGGCCAGAATCGGAGCGAGTGTCATTTTGCTACAAGGAACGATA 322

Qy        106 -------ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGlyAs 123

Db        321 TAGCCTTCAGCAGGAGAAGAACGATCCAGACAAGCCCAATTGCCACGCCAAGGATGGCAA 262
```

```
Qy        123 nProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPheTh 143
              ||||||||||||||||||||||    ||||||:::   |||   :::||||||   :::
Db        261 TCCTTTTGGTCCCTTTTGGGACACTTTTCACATAGACTTTGTGCGGTCAGAGTTCTATGC 202

Qy        143 rGlyIleSerPheSerAlaSerTyrArgGluGln---------TrpSerGlnArgPheSe 160
              :::   |||         :::     :::          |||      ::::::
Db        201 GCCACTTCATTTTGATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCAAATATCC 142

Qy        160 rProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGl 180
              :::   :::||||||||||||      ||||||||||||   |||||||||
Db        141 TGCAGAATCATATCCCGTACTCGCGTTCACCGGAGCTCCGGCTAGTTTTCCTGTTCAGCT 82

Qy        180 uGluHisArgProLeuGlnLysTyrMetValTrpSer-AspGluMetValLysThrGlyG 200
               |||:::      ||||||:::|||:::   |||||| :::
Db         81 AGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGT--------------- 38

Qy        200 luAlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlyS 220
                                        :::|||   ||||||||||||   |
Db         37 ---------------------------------ATAGGGAAGCATCTAAGGATTTCAT 13

Qy        220 erAsp 221
              ||:::
Db         12 CCGAG 8
```

RESULT 13
US-09-764-891-6032
; Sequence 6032, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6032
; LENGTH: 10331
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6032

Alignment Scores:			
Pred. No.:	5.16e-41	Length:	10331
Score:	424.00	Matches:	83
Percent Similarity:	98.81%	Conservative:	0
Best Local Similarity:	98.81%	Mismatches:	1
Query Match:	21.42%	Indels:	0
DB:	10	Gaps:	0

US-09-774-954-2 (1-365) x US-09-764-891-6032 (1-10331)

```
Qy        223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2491 AAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCG 2550

Qy        243 GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro 262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2551 CAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCT 2610

Qy        263 AspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSer 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2611 GACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCG 2670

Qy        283 ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLysGly 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2671 GTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGG 2730

Qy        303 LysValLysVal 306
              ||||||   |||
Db       2731 AAGGTATGTGTG 2742
```

RESULT 14
US-09-918-995-19437
; Sequence 19437, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19437
; LENGTH: 479
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-19437

Alignment Scores:			
Pred. No.:	4.16e-39	Length:	479
Score:	393.00	Matches:	76
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	19.86%	Indels:	0
DB:	10	Gaps:	0

US-09-774-954-2 (1-365) x US-09-918-995-19437 (1-479)

```
Qy        290 GluSerTyrValProGluLeuGlnGlnLeuPheLysGlyLysValLysValValSerLeu 309
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 GAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGGAAGGTGAAGGTGGTGAGCCTG 90

Qy        310 LysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAspHisPheIleGly 329
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 AAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGCCAAGCCGACCACTTTATTGGC 150

Qy        330 AsnCysValSerSerPheThrAlaPheValLysArgGluArgAspLeuGlnGlyArgPro 349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 AACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAGCGGGACCTCCAGGGGAGGCCG 210

Qy        350 SerSerPhePheGlyMetAspArgProProLysLeuArgAspGluPhe 365
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 TCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGGGACGAGTTC 258
```

RESULT 15
US-11-097-143-18169/c
; Sequence 18169, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831

```
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18169
;   LENGTH: 3567
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-18169

Alignment Scores:
Pred. No.:              1.14e-37        Length:         3567
Score:                  392.00          Matches:        88
Percent Similarity:     57.80%          Conservative:   38
Best Local Similarity:  40.37%          Mismatches:     76
Query Match:            19.81%          Indels:         16
DB:                     26              Gaps:           6

US-09-774-954-2 (1-365) x US-11-097-143-18169 (1-3567)

Qy        155 TrpSerGlnArgPheSerProLysGluHisProValLeuAlaLeuProGlyAlaProAla 174
              |||      ::::::     :::   :::||||||||||||     ||||||||||||
Db       3567 TGGCAGACCAAATATCCTGCAGAATCATATCCCGTACTCGCGTTCACCGGAGCTCCGGCT 3508

Qy        175 GlnPheProValLeuGluGluHisArgProLeuGlnLysTyrMetValTrpSerAspGlu 194
                |||||||||      |||:::      ||||||:::|||:::   ||||||
Db       3507 AGTTTTCCTGTTCAGCTAGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGG 3448

Qy        195 MetValLysThrGlyGluAlaGlnIleHisAlaHisLeuValArg---ProTyrValGly 213
                :::      :::     |||          |||   |||     ::::::|||
Db       3447 TATAGGGAAGCATCTAAGGATTTCATCCGAGAGCAGTTGCCTCGGGGTGCCTTTTTGGGC 3388

Qy        214 IleHisLeuArgIleGlySerAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThr 233
              ||||||||||||   |||   ||||||      ||||||      :::||||||
Db       3387 ATTCATCTGCGCAACGGTATCGATTGGGTGAGAGCCTGTGAGCACGTCAAGGAT------ 3334

Qy        234 AlaGlySerHisPheMetAlaSerProGlnCysValGlyTyrSerArgSerThrAlaAla 253
                  |||      ||||||||||||||:::||||||                  |||
Db       3333 ---AGCCAGCATCTGTTTGCCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTGCA 3277

Qy        254 ProLeuThrMetThrMetCysLeuProAspLeuLysGluIleGlnArgAlaValLysLeu 273
               |||        :::|||:::|||      :::   |||   |||   :::|||
Db       3276 ---CTCTACCCGGAGCTCTGCATGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGA 3220

Qy        274 TrpValArgSerLeu--------------------AspAlaGlnSerValTyrValAla 286
                :::::::::::                    :::   :::||||||:::||||||
Db       3219 ACCATTAAGAACGTGCGCCAAACTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCG 3160

Qy        287 ThrAspSerGluSerTyrValProGluLeuGln---GlnLeuPheLysGlyLysValLys 305
              :::||||||         :::   ||||||         |||   :::      :::
Db       3159 TCAGACTCCAATCACATGATTGGTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGT 3100

Qy        306 ValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAsp 325
              |||      |||      :::         :::|||||||   ||||||||||||:::::
Db       3099 GTGCACAAGCTGCCGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAAC 3040

Qy        326 HisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArgGluArgAspLeu 345
              ||||||||||||||||:::||||||:::::||||||   ||||||||||||||:::
Db       3039 CACTTTATCGGCAACTGTATATCCTCTTACTCGGCATTCGAAAAAAGGGAACGAGATGTG 2980

Qy        346 GlnGlyArgProSerSerPhePheGlyMetAspArgProProLysLeuArgAsp 363
                |||   ||||||   |||:::|||            ||||||   :::|||
Db       2979 CACGGTTTTCCATCGTACTTCTGGGGATTC---------CCCAAGGAAAAGGAT 2935


Search completed: October 26, 2005, 15:59:38
Job time : 1594.35 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 05:59:38 ; Search time 6375.13 Seconds
(without alignments)
2179.320 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....QGRPSSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105432_6406/app_query.fasta_1.917
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954_@CGN_1_1_5334_@runat_25102005_105432_6406 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1979	100.0	1555	3	CR617950	CR617950 full-leng
2	1832	92.6	3611	3	AK081059	AK081059 Mus muscu
3	1829	92.4	2616	3	AK044629	AK044629 Mus muscu
4	1525	77.1	1141	5	BX353274	BX353274 BX353274
5	1394	70.4	1091	5	BM909055	BM909055 AGENCOURT
6	1356	68.5	801	5	BP144264	BP144264 BP144264
7	1355	68.5	947	4	BI410783	BI410783 602963556
8	1350.5	68.2	889	4	BI524031	BI524031 603052202
9	1339	67.7	988	5	BQ964179	BQ964179 AGENCOURT
10	1329	67.2	841	5	BX345989	BX345989 BX345989
11	1316	66.5	915	5	BQ936634	BQ936634 AGENCOURT
12	1304	65.9	898	5	BX449172	BX449172 BX449172
13	1304	65.9	942	7	CO774833	CO774833 ILLUMIGEN
14	1272	64.3	719	1	AL135434	AL135434 DKFZp762O
15	1237	62.5	808	4	BG747760	BG747760 602705255
16	1196	60.4	874	4	BI757540	BI757540 603029693
17	1178	59.5	921	1	AJ454995	AJ454995 AJ454995
18	1176	59.4	928	7	CF239348	CF239348 AGENCOURT
19	1167	59.0	919	5	BU155142	BU155142 AGENCOURT
20	1164.5	58.8	809	4	BG473023	BG473023 602514976
21	1164	58.8	927	6	CA974668	CA974668 AGENCOURT
22	1152	58.2	705	7	CN348925	CN348925 170005328
23	1126	56.9	753	4	BG424731	BG424731 602453575
24	1118	56.5	760	5	BU055535	BU055535 UI-M-FO0-
25	1099	55.5	775	5	BU485326	BU485326 603843502
26	1087	54.9	689	4	BI559635	BI559635 603252494
27	1073	54.2	763	7	CK366990	CK366990 AGENCOURT
28	1044	52.8	819	7	CF219216	CF219216 AGENCOURT
29	1027	51.9	590	4	BG894786	BG894786 355525 MA
30	1021	51.6	666	5	BP459416	BP459416 BP459416
31	1019	51.5	736	5	BP457865	BP457865 BP457865
32	1001	50.6	561	4	BM798261	BM798261 K-EST0081
33	987	49.9	644	7	CN459072	CN459072 UI-M-HN0-
34	976	49.3	731	1	AJ453431	AJ453431 AJ453431
35	976	49.3	4147	3	AK053648	AK053648 Mus muscu
36	964	48.7	695	2	BB617910	BB617910 BB617910
37	963	48.7	664	6	BY728337	BY728337 BY728337
38	961	48.6	743	5	BP686346	BP686346 BP686346
39	957	48.4	624	7	CN348922	CN348922 170004249
40	952	48.1	641	2	BB611933	BB611933 BB611933
41	944	47.7	661	2	BB642654	BB642654 BB642654
42	943	47.7	699	1	AU137536	AU137536 AU137536
43	941	47.5	662	4	BM426089	BM426089 pgf2n.pk0
44	940	47.5	746	5	BP699469	BP699469 BP699469
45	936	47.3	545	7	CN348928	CN348928 170005313

ALIGNMENTS

RESULT 1
CR617950
LOCUS CR617950 1555 bp mRNA linear HTC 21-JUL-2004
DEFINITION full-length cDNA clone CS0DJ008YF12 of T cells (Jurkat cell line) Cot 10-normalized of Homo sapiens (human).
ACCESSION CR617950
VERSION CR617950.1 GI:50498757
KEYWORDS HTC; CNSLT_cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1555)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue
REFERENCE 2 (bases 1 to 1555)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
FEATURES Location/Qualifiers
source 1. .1555
/organism="Homo sapiens"

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ008YF12"
                     /tissue_type="T cells (Jurkat cell line) Cot
                     10-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

Alignment Scores:
Pred. No.:                3.53e-204       Length:          1555
Score:                    1979.00         Matches:         365
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       3               Gaps:            0

US-09-774-954-2 (1-365) x CR617950 (1-1555)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        109 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 168

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        169 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 228

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 288

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        289 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 348

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        349 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 408

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        409 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 468

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        469 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 528

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        529 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 588

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        589 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 648

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        649 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 708

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        709 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 768

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        769 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 828

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        829 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 888

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        889 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 948

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        949 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1008

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1009 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1068

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1069 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1128

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1129 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1188

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1189 CTGCGGGACGAGTTC 1203

RESULT 2
AK081059
LOCUS       AK081059               3611 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length
            enriched library, clone:B930076G19 product:protein
            O-fucosyltransferase 1, full insert sequence.
ACCESSION   AK081059
VERSION     AK081059.1  GI:26099647
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 3611)
```

```
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES              Location/Qualifiers
     source           1. .3611
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="FANTOM_DB:B930076G19"
                      /db_xref="taxon:10090"
                      /clone="B930076G19"
                      /tissue_type="cerebellum"
                      /clone_lib="RIKEN full-length enriched mouse cDNA library"
                      /dev_stage="10 days neonate"
     misc_feature     11. .1194
                      /note="protein O-fucosyltransferase 1 (MGD|MGI:2153207,
                      GB|NM_080463, evidence: BLASTN, 99%, match=1290)
                      putative"
ORIGIN

Alignment Scores:
Pred. No.:                1.41e-187      Length:          3611
Score:                    1832.00        Matches:         336
Percent Similarity:       96.14%         Conservative:    13
Best Local Similarity:    92.56%         Mismatches:      14
Query Match:              92.57%         Indels:          0
DB:                       3              Gaps:            0

US-09-774-954-2 (1-365) x AK081059 (1-3611)

Qy          3 AlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGly 22
              |||||||||||||||   |||||||||||||||||||||||||||||||||||||||||||
Db        103 GCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGCTTTGGG 162

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        163 AACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGCACCTTG 222

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisVal 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        223 GCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTCCATGTG 282

Qy         63 SerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeu 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db        283 TCCTACCAAAAGTACTTCAAACTGGAGCCTCTCCAAGCCTACCATCGGGTTGTCAGCCTG 342

Qy         83 GluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyr 102
              |||||||||||||||  |||||||||:::||||||||||||||||||||||||||||||

Db        343 GAGGACTTCATGGAAAATCTGGCACCCTCCCACTGGCCCCCTGAGAAGCGAGTGGCATAC 402

Qy        103 CysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGly 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 TGCTTTGAGGTGGCAGCCCAGCGAAGTCCTGATAAGAAGACATGTCCCATGAAGGAAGGA 462

Qy        123 AsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPhe 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 AATCCTTTTGGGCCATTCTGGGACCAGTTTCATGTGAGTTTCAATAAGTCAGAACTGTTC 522

Qy        143 ThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerProLys 162
              ||||||||||||||||||||||||||:::|||||||||:::||||||||||      |||
Db        523 ACAGGCATTTCCTTCAGCGCCTCCTACAAAGAACAATGGACCCAGAGATTTCCTGCAAAA 582

Qy        163 GluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGluHis 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 GAGCATCCTGTGCTCGCACTGCCTGGGGCCCCAGCACAGTTCCCTGTCCTGGAGGAACAC 642

Qy        183 ArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAlaGln 202
              |||  ||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db        643 AGGGAGCTCCAGAAGTACATGGTGTGGTCAGATGAGATGGTGAGGACGGGAGAGGCCCTG 702

Qy        203 IleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTrp 222
              |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 ATCAGTGCCCACCTCGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCCGACTGG 762

Qy        223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 AAGAATGCCTGTGCCATGCTGAAGGATGGAACTGCAGGGTCACACTTCATGGCTTCCCCT 822

Qy        243 GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro 262
              ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db        823 CAGTGTGTGGGCTATAGCCGCAGCACAGCGACCCCTCTCACCATGACCATGTGCCTCCCT 882

Qy        263 AspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSer 282
              |||||||||||||||||||||||||||   ||||||||||:::|||:::|||:::|||
Db        883 GACCTGAAGGAAATCCAGCGGGCTGTGACGCTTTGGGTGAGAGCACTGAATGCCAGATCG 942

Qy        283 ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLysGly 302
              |||||||:::|||||||||||||||||||||||||   |||:::|||||||||||||||
Db        943 GTCTACATCGCCACAGACTCTGAGAGCTACGTGTCAGAGATCCAGCAGCTCTTCAAAGAC 1002

Qy        303 LysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGly 322
              ||||||:::|||||||||||||||||||||||||||||:::|||||||||||||||||||
Db       1003 AAGGTGAGGGTGGTGAGCCTGAAACCCGAGGTGGCCCAGATCGACCTGTACATCCTCGGC 1062

Qy        323 GlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArgGlu 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 CAGGCTGACCACTTCATTGGAAACTGTGTCTCCTCGTTCACTGCCTTCGTGAAGCGGGAG 1122

Qy        343 ArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLysLeuArg 362
              |||||||||   ||||||   ||||||||||||||||||||||||||   :::||||||
Db       1123 CGGGACCTCCATGGGAGGCAGTCGTCCTTCTTTGGCATGGACAGACCCTCCCAGCTTCGG 1182

Qy        363 AspGluPhe 365
              |||||||||
Db       1183 GATGAATTT 1191

RESULT 3
AK044629
LOCUS       AK044629               2616 bp    mRNA    linear   HTC 03-APR-2004
DEFINITION  Mus musculus adult retina cDNA, RIKEN full-length enriched library,
            clone:A930028F21 product:protein O-fucosyltransferase 1, full
            insert sequence.
ACCESSION   AK044629
VERSION     AK044629.1  GI:26336652
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
```

```
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2616)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Retina RNA was provided by Dr. Stefano Gustincich (Department of
            Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA
            02115, USA) whose assistance is gratefully acknowledged. Please
            visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES             Location/Qualifiers
     source          1. .2616
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:A930028F21"
                     /db_xref="taxon:10090"
                     /clone="A930028F21"
                     /tissue_type="retina"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             34. .1215
                     /note="unnamed protein product; protein
                     O-fucosyltransferase 1 (MGD|MGI:2153207, GB|NM_080463,
                     evidence: BLASTN, 99%, match=1290)
                     putative"
                     /codon_start=1
                     /protein_id="BAC32009.1"
                     /db_xref="GI:26336653"
                     /translation="MGAAAWAPPHLLLRASFLLLLLLLPLRGRSAGSWDLAGYLLYCP
                     CMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQ
                     AYHRVVSLEDFMENLAPSHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQF
                     HVSFNKSELFTGISFSASYKEQWTQRFPAKEHPVLALPGAPAQFPVLEEHRELQKYMV
                     WSDEMVRTGEALISSHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYS
                     RSTATPLTMTMCLPDLKEIQRAVTLWVRALNARSVYIATDSESYVSEIQQLFKDKVRV
                     VSLKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLHGRQSSFFGMDRPSQLRDE
                     F"
ORIGIN
```

Alignment Scores:

Pred. No.:	1.79e-187	Length:	2616
Score:	1829.00	Matches:	335
Percent Similarity:	96.14%	Conservative:	14
Best Local Similarity:	92.29%	Mismatches:	14
Query Match:	92.42%	Indels:	0
DB:	3	Gaps:	0

US-09-774-954-2 (1-365) x AK044629 (1-2616)

```
Qy          3 AlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGly 22
              ||||||||||||||    ||||||||||||||||||||||||||||||||||||||||||
Db        124 GCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGCTTTGGG 183

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        184 AACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGCACCTTG 243

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisVal 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 GCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTCCATGTG 303

Qy         63 SerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeu 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        304 TCCTACCAAAAGTACTTCAAACTGGAGCCTCTCCAAGCCTACCATCGGGTTGTCAGCCTG 363

Qy         83 GluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyr 102
              ||||||||||||||    |||||||||:::||||||||||||||||||||||||||||||
Db        364 GAGGACTTCATGGAAAATCTGGCACCCTCCCACTGGCCCCCTGAGAAGCGAGTGGCATAC 423

Qy        103 CysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGly 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        424 TGCTTTGAGGTGGCAGCCCAGCGAAGTCCTGATAAGAAGACATGTCCCATGAAGGAAGGA 483

Qy        123 AsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPhe 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        484 AATCCTTTTGGGCCATTCTGGGACCAGTTTCATGTGAGTTTCAATAAGTCAGAACTGTTC 543

Qy        143 ThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerProLys 162
              ||||||||||||||||||||||||||:::|||||||||:::|||||||||      |||
Db        544 ACAGGCATTTCCTTCAGCGCCTCCTACAAAGAACAATGGACCCAGAGATTTCCTGCAAAA 603

Qy        163 GluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGluHis 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        604 GAGCATCCTGTGCTCGCACTGCCTGGGGCCCCAGCACAGTTCCCTGTCCTGGAGGAACAC 663

Qy        183 ArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAlaGln 202
              |||   ||||||||||||||||||||||||||||||||||:::||||||||||||
Db        664 AGGGAGCTCCAGAAGTACATGGTGTGGTCAGATGAGATGGTGAGGACGGGAGAGGCCCTG 723
```

```
Qy        203 IleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTrp 222
              |||   :::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        724 ATCAGTTCCCACCTCGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCCGACTGG 783

Qy        223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        784 AAGAATGCCTGTGCCATGCTGAAGGATGGAACTGCAGGGTCACACTTCATGGCTTCCCCT 843

Qy        243 GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro 262
              |||||||||||||||||||||||||||||||   |||||||||||||||||||||||||
Db        844 CAGTGTGTGGGCTATAGCCGCAGCACAGCGACCCCTCTCACCATGACCATGTGCCTCCCT 903

Qy        263 AspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSer 282
              |||||||||||||||||||||||||   ||||||||||:::|||:::|||:::|||
Db        904 GACCTGAAGGAAATCCAGCGGGCTGTGACGCTTTGGGTGAGAGCACTGAATGCCAGATCG 963

Qy        283 ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLysGly 302
              ||||||:::|||||||||||||||||||||||||   |||:::|||||||||||||||
Db        964 GTCTACATCGCCACAGACTCTGAGAGCTACGTGTCAGAGATCCAGCAGCTCTTCAAAGAC 1023

Qy        303 LysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGly 322
              ||||||:::|||||||||||||||||||||||||||:::|||||||||||||||||||
Db       1024 AAGGTGAGGGTGGTGAGCCTGAAACCCGAGGTGGCCCAGATCGACCTGTACATCCTCGGC 1083

Qy        323 GlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArgGlu 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1084 CAGGCTGACCACTTCATTGGAAACTGTGTCTCCTCGTTCACTGCCTTCGTGAAGCGGGAG 1143

Qy        343 ArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLysLeuArg 362
              |||||||||   ||||||   |||||||||||||||||||||||||||   :::||||||
Db       1144 CGGGACCTCCATGGGAGGCAGTCGTCCTTCTTTGGCATGGACAGACCCTCCCAGCTTCGG 1203

Qy        363 AspGluPhe 365
              |||||||||
Db       1204 GATGAATTT 1212
```

RESULT 4
BX353274
LOCUS BX353274 1141 bp mRNA linear EST 08-APR-2004
DEFINITION BX353274 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DC003YJ11 5-PRIME, mRNA sequence.
ACCESSION BX353274
VERSION BX353274.2 GI:46290116
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1141)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On May 5, 2003 this sequence version replaced gi:30371746.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9980.f
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0DC003CE06QP1&c=9980.f.

FEATURES Location/Qualifiers
source 1. .1141
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC003YJ11"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:

Pred. No.:	6.47e-155	Length:	1141
Score:	1525.00	Matches:	311
Percent Similarity:	87.19%	Conservative:	2
Best Local Similarity:	86.63%	Mismatches:	40
Query Match:	77.06%	Indels:	10
DB:	5	Gaps:	4

US-09-774-954-2 (1-365) x BX353274 (1-1141)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db         86 ATGCCCGCRGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCMCCTGCATGGGGCGC 145

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||   |||||||||   ||||||||||||||||||||||||||||||||||||||
Db        146 TTTGGGAAMCAGGCCGATMACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCWAAAMMGT 205

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||   |||   ||||||||||||||||
Db        206 ACCTTGGCTGTACCTCCTTGGATTGAGTACCAGCATCAMAAGCMTCCTTTCACCAACCTC 265

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||   |||   |||   ||||||||||||   |||||||||
Db        266 CATGTGTCCTACCAGAAGTACTTMAARMWGGAGCACCTMCAGGCTTACMATCGGGTCATC 325

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 385

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||   |||   ||||||||||||||||||||
Db        386 GCATACTGCTTTGAGGTGGCAGCMCAGCGAAG-CCAGAWAAGAAGACGTGCCCCATGAAG 444

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        445 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTATCAACAAGTCGGAG 504

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        505 CTTTTTACAGGCATTTMCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 564

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 624

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        625 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAR 684

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        685 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 744

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        745 GACTGGAAGAACGCCTGTGTCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 804

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 864

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrp-ValArgSerLeuAspAl 280
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        865 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTKAAGCTCTGGGGTGAGGTCGCTGGATGC 924
```

```
Qy        280 aGlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPh 300
              |||||||||||||||||||||||||||||||||||||||||||    |||||||||||||
Db        925 CCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTTCCTKAGCTCCAACAGCTCTT 984

Qy        300 eLysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIl 320
              ||||||||||||||||||||    ||||||  |||     ||||||||
Db        985 CAAAGGGAAGGTGAAGGTGGTGAG-CTKAAGCYTGAGGK-GSCCAGGTCGACTGT---AM 1039

Qy        320 eLeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLy 340
                      :::  ||||||||||||||     |||   :::       |||||||
Db       1040 ATCCTCGGCAAGCGRACACTTTATTGGCAATKT-KTTTCTCYTTAT---TGCTTTGTKAA 1095

Qy        340 sArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgPro 358
              |      ||     |||||            ||||||     ||||||
Db       1096 G---GGGAGGGGACMYCAGGGG------RGCGTTTTTTTTYGGMATGAAGGCCC 1141

RESULT 5
BM909055
LOCUS       BM909055                1091 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  AGENCOURT_6611915 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5431669
            5', mRNA sequence.
ACCESSION   BM909055
VERSION     BM909055.1  GI:19359434
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1091)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1904  row: e  column: 14
            High quality sequence stop: 629.
FEATURES             Location/Qualifiers
     source          1..1091
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5431669"
                     /tissue_type="astrocytoma grade IV, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_98"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:              1.14e-140      Length:         1091
Score:                  1394.00        Matches:        290
Percent Similarity:     79.78%         Conservative:   6
Best Local Similarity:  78.17%         Mismatches:     20
Query Match:            70.44%         Indels:         55
DB:                     5              Gaps:           5

US-09-774-954-2 (1-365) x BM909055 (1-1091)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATG------ 130

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40

Db        130 ------------------------------------------------------------ 130

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60

Db        131 ---------------------------------------------------------GCT 133

Qy         61 -HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIl 80
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 CCATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCAT 193

Qy         80 eSerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVa 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 CAGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGT 253

Qy        100 lAlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLy 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        254 GGCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAA 313

Qy        120 sGluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGl 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        314 GGAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGA 373

Qy        140 uLeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSe 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        374 GCTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTC 433

Qy        160 rProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGl 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        434 TCCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGA 493

Qy        180 uGluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGl 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        494 GGAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGA 553

Qy        200 uAlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySe 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        554 GGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTC 613

Qy        220 rAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAl 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        614 TGACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGC 673

Qy        240 aSerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCy 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        674 CTCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTG 733

Qy        260 sLeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAl 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        734 CCTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGC 793

Qy        280 aGlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPh 300
              |||||||||||||||||||||||||||||||||||||||||||||       ||||||||
Db        794 CCAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAAGCTCCACAGCTCTT 853

Qy        300 e--LysGlyLysValLysValVal-SerLeuLysProGlu-ValAlaGlnValAspLeuT 319
              |  ||||||:::   |||||||| :::||||||||       ||||||||   |||
Db        854 TAAAAGGGAAGGGGGAAGGTGGTGGAACCTGAAACCTGTAGGGGGCCCAGGTCCACCTGT 913

Qy        319 yrIleLeuGlyGlnAlaAspHisPhe-IleGlyAsnCysValSerSerPheThrAla--- 337
                 ||||||   ||||||:::   |||  |||||||||||
Db        914 AAATTCTTCGGCAAGCCGAACCCTTTAATTGGGAACTGTGGCCCCCTCCTTCCCGGGCCC 973

Qy        338 PheValLysArgGluArgAsp---LeuGlnGlyArg---------ProSerSerPhePhe 353
              |||   |||    :::||||||   |||:::|||           ||||||     |||
Db        974 TTTGGAAAAGGGAAACGGACCCTCTCAGGGGAGGGGCGGCTTTTCCTTCT------TTC 1027

Qy        354 GlyMet-AspArgProPro 359
                 |||  :::|||||||||
```

```
Db       1028 GCAATGGGAAAGGCCCCCC 1046

RESULT 6
BP144264
LOCUS       BP144264                  801 bp    mRNA    linear   EST 30-DEC-2003
DEFINITION  BP144264 full-length enriched swine cDNA library, adult ovary Sus
            scrofa cDNA clone OVR010034G09 5', mRNA sequence.
ACCESSION   BP144264
VERSION     BP144264.1  GI:40393735
KEYWORDS    EST.
SOURCE      Sus scrofa (pig)
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 801)
  AUTHORS   Uenishi,H., Eguchi,T., Suzuki,K., Sawazaki,T., Toki,D., Shinkai,H.,
            Okumura,N., Hamasima,N. and Awata,T.
  TITLE     PEDE (Pig EST Data Explorer): construction of a database for ESTs
            derived from porcine full-length cDNA libraries
  JOURNAL   Nucleic Acids Res. 32 (1), D484-D488 (2004)
COMMENT     Contact: Hirohide Uenishi
            Animal Genome Laboratory, Genome Research Department
            National Institute of Agrobiological Sciences
            2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan
            Tel: +81-29-838-8627
            Fax: +81-29-838-8627
            Email: huenishi@affrc.go.jp
            EST project with full-length enriched cDNA libraries carried out in
            Animal Genome Research Program (Japan) by National Institute of
            Agrobiological Sciences and STAFF-Institute
            Single pass sequencing of clones derived from oligo-capped cDNA
            library
            Vector sequences were eliminated by RepeatMasker version 2002/07/13
            and crossmatch version 0.990319
            Low quality bases were trimmed based on the quality values.
FEATURES             Location/Qualifiers
     source          1. .801
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /clone="OVR010034G09"
                     /tissue_type="ovary"
                     /dev_stage="adult"
                     /clone_lib="full-length enriched swine cDNA library, adult
                     ovary"
ORIGIN

Alignment Scores:
Pred. No.:              9.66e-137       Length:         801
Score:                  1356.00         Matches:        246
Percent Similarity:     97.37%          Conservative:   13
Best Local Similarity:  92.48%          Mismatches:     7
Query Match:            68.52%          Indels:         0
DB:                     5               Gaps:           0

US-09-774-954-2 (1-365) x BP144264 (1-801)

Qy         66 LysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeuGluAspPhe 85
              ||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db          3 AAGTACTTCAAGCTGGAGCCCCTCCAGGTCTACCACCGGGTCATCAGCCTGGAGGACTTC 62

Qy         86 MetGluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyrCysPheGlu 105
              ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db         63 ATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAGGCGGGTGGCATACTGTTTTGAG 122

Qy        106 ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGlyAsnProPhe 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 GTGGCAGCACAGCGAAGCCCAGATAAGAAGACATGCCCCATGAAGGAAGGAAACCCCTTT 182

Qy        126 GlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPheThrGlyIle 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        183 GGCCCATTTTGGGATCAGTTTCATGTGAGTTTCAACAAGTCAGAACTTTTTGCAGGCATT 242

Qy        146 SerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerProLysGluHisPro 165
              ||||||||||||||||||:::::||||||   ||||||||||||||||||||||||||||
Db        243 TCCTTCAGTGCCTCCTACAAAGACCAGTGGATCCAAAGATTTTCTCCAAAGGAACATCCA 302

Qy        166 ValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGluHisArgProLeu 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 GTGCTTGCCCTGCCAGGGGCCCCGGCTCAGTTCCCCGTCCTGGAGGAACATAGGCCACTT 362

Qy        186 GlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAlaGlnIleHisAla 205
              ||||||||||||||||||||||||||||||:::||||||||||||||||||   |||
Db        363 CAGAAATACATGGTGTGGTCAGACGAGATGGTGAGGACAGGGGAGGCCCAGATCCGTGCC 422

Qy        206 HisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTrpLysAsnAla 225
              ||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 CACCTCATCCGGCCCTACGTGGGCATTCACCTGCGCATTGGCTCTGACTGGAAGAACGCG 482

Qy        226 CysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerProGlnCysVal 245
              ||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        483 TGCGCCATGCTGAAGGACGGCACTGCCGGCGCTCACTTCATGGCCTCCCCGCAGTGCGTG 542

Qy        246 GlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuProAspLeuLys 265
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db        543 GGTTACAGCCGCCACACCGCCGCCCCGCTCACCATGACCATGTGCCTCCCCGACCTGAAG 602

Qy        266 GluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSerValTyrVal 285
              ||||||:::||||||:::||||||||||||   :::|||   |||||||||||||:::
Db        603 GAGATCAGGCGCGCCCTGAAGCTCTGGGTGACGGCGCTGAGCGCCCAGTCCGTCTACATC 662

Qy        286 AlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLysGlyLysValLys 305
              ||||||||||||||||||:::||||||:::||||||||||||||||||||||||||||
Db        663 GCCACGGACTCTGAGAGTTACCTGCCGGAGATCCAGCAGCTCTTCAAAGGAAAGGTGAAG 722

Qy        306 ValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAsp 325
              ||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        723 GTGGTGAGTCTGAAACCTGAGGTGGCCCAGATCGACCTGTACATCCTTGGCCAGGCTGAC 782

Qy        326 HisPheIleGlyAsnCys 331
              ||||||||||||||||||
Db        783 CACTTTATTGGCAACTGT 800

RESULT 7
BI410783
LOCUS       BI410783                  947 bp    mRNA    linear   EST 14-AUG-2001
DEFINITION  602963556F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5119332 5',
            mRNA sequence.
ACCESSION   BI410783
VERSION     BI410783.1  GI:15171706
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 947)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
            Bonaldo, Ph.D.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by:Incyte Genomics, Inc.
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11290  row: o  column: 13
            High quality sequence start: 21
            High quality sequence stop: 919.
FEATURES             Location/Qualifiers
     source          1. .947
                     /organism="Mus musculus"
                     /mol_type="mRNA"
```

```
                  /strain="CZECH II"
                  /db_xref="taxon:10090"
                  /clone="IMAGE:5119332"
                  /tissue_type="pooled lung tumors"
                  /lab_host="DH10B (phage-resistant)"
                  /clone_lib="NCI_CGAP_Lu33"
                  /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
                  modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
                  strand cDNA was prepared from mRNA obtained from pooled
                  lung tumors with a Not I - oligo(dT) primer [5'
                  TGTTACCAATCTGAAGTGGGAGCGGCCGCCTCTGTTTTTTTTTTTTTTTT 3']. 
                  Double-stranded cDNA was ligated to Eco RI adaptors
                  (Pharmacia),  digested with Not I and cloned into the Not
                  I and Eco RI sites of  the modified pT7T3 vector. Library
                  went through one round of normalization, and was
                  constructed  by Bento Soares and M. Fatima Bonaldo. "
ORIGIN

Alignment Scores:
Pred. No.:              1.62e-136      Length:         947
Score:                  1355.00        Matches:        268
Percent Similarity:     94.18%         Conservative:   7
Best Local Similarity:  91.78%         Mismatches:     16
Query Match:            68.47%         Indels:         7
DB:                     4              Gaps:           0

US-09-774-954-2 (1-365) x BI410783 (1-947)

Qy          3 AlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGly 22
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db         74 GCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGCTTTGGG 133

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        134 AACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGCACCTTG 193

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisVal 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        194 GCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTCCATGTG 253

Qy         63 SerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeu 82
              ||||||||||||||||||||||||||||||||||||||||||||||||:::|||||||
Db        254 TCCTACCAAAAGTACTTCAAACTGGAGCCTCTCCAAGCCTACCATCGGGTTGTCAGCCTG 313

Qy         83 GluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyr 102
              ||||||||||||||   |||||||||:||||||||||||||||||||||||||||||||
Db        314 GAGGACTTCATGGAAAATCTGGCACCCTCCCACTGGCCCCCTGAGAAGCGAGTGGCATAC 373

Qy        103 CysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGly 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        374 TGCTTTGAGGTGGCAGCCCAGCGAAGTCCTGATAAGAAGACATGTCCCATGAAGGAAGGA 433

Qy        123 AsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPhe 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        434 AATCCTTTTGGGCCATTCTGGGACCAGTTTCATGTGAGTTTCAATAAGTCAGAACTGTTC 493

Qy        143 ThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerProLys 162
              |||||||||||||||||||||||:::|||||||||:::|||||||||       |||
Db        494 ACAGGCATTTCCTTCAGCGCCTCCTACAAAGAACAATGGACCCAGAGATTTCCTGCAAAA 553

Qy        163 GluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProVal-LeuGluGluHi 182
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        554 GAGCATCCTGTGCTCGCACTGCCTGGGGCCCCAGCACAGTTCCCTGTCCCTGGAGGAACA 613

Qy        182 sArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAlaGl 202
              ||||   ||||||||||||||||||||||||||||||||||:::||||||||||||||
Db        614 CAGGGAGCTCCAGAAGTACATGGTGTGGTCAGATGAGATGGTGAGGACGGGAGAGGCCCT 673

Qy        202 nIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTr 222
               |||   |||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        674 GATCAGTGCCCACCTCGTCCGGCCCTATGTGGGCATTCATCTGCGCAT-GGCTCCGACTG 732

Qy        222 pLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPr 242
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        733 GAAGAATGCCTGTGCCATGCTGAAGGATGGAACTGCAGGCTCACACTTCATGGCTTCCCC 792

Qy        242 oGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPr 262
              |||||||||||||||||||||||||||||      ||||||   ||||||||||||||||
Db        793 TCAGTGTGTGGGCTATAGCCGCAGCACAAGGAC-CCTCTCAC-ATGACCATGTGCCTCCC 850

Qy        262 oAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSe 282
              ||||||||||||   ||||||||||||   ||||||||||:::|||      |||||
Db        851 TGACCTGAAGGAAAT-CAGCGGGCTGTGACGCTTTGGGTGAGAGCATTGA--TGCCAATC 907

Qy        282 rValTyrValAlaThrAspSerGluSerTyrVal 293
              |||||||:::   ||||||||||||||||||||
Db        908 GGTCTACATCGC-ACAGACTCTGAGAGCTACGTG 940

RESULT 8
BI524031
LOCUS       BI524031                889 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  603052202F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201616 5',
            mRNA sequence.
ACCESSION   BI524031
VERSION     BI524031.1  GI:15348823
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 889)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11505  row: d  column: 01
            High quality sequence stop: 880.
FEATURES             Location/Qualifiers
     source          1. .889
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5201616"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_122"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:              4.53e-136      Length:         889
Score:                  1350.50        Matches:        257
Percent Similarity:     95.57%         Conservative:   2
Best Local Similarity:  94.83%         Mismatches:     8
Query Match:            68.24%         Indels:         6
DB:                     4              Gaps:           1

US-09-774-954-2 (1-365) x BI524031 (1-889)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           87 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 146

Qy           21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          147 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 206

Qy           41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          207 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 266

Qy           61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          267 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 326

Qy           81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          327 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 386

Qy          101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          387 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 446

Qy          121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          447 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 506

Qy          141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          507 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 566

Qy          161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          567 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 626

Qy          181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          627 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 686

Qy          201 AlaGlnIleHisAlaHisLeuValArgPro-TyrValGlyIleHisLeuArgIleGlySe 220
                |||||||||||||||    ||||||||||||| ||||||||||||||||||||||||||||
Db          687 GCCCAGATTCATGCCCAGCTGGTCCGGCCACTATGTGGGCATTCATCTGCGCATTGGCTC 746

Qy          220 rAspTrp-LysAsnAlaCysAlaMetLeuLysAsp-GlyThrAlaGlySerHisPheMet 239
                |||||| |||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db          747 TGACTGGGAAGAACGCCTGTGCCATGCTGAAGGACGGGAACTGCAGGCTCGCACTTCATG 806

Qy          240 AlaSerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMet 259
                ||||||||||||||||||||||||||||  :::          |||    ||||||||
Db          807 GCCTCTCCGCAGTGTGTGGGCTACAGCGCAGCACAGCGGCCCCCT---CAGATGACTATG 863

Qy          260 CysLeuProAspLeuLysGluIleGln 268
                |||  ||||||   |||||||||:::
Db          864 TGCTG-CCTGACTG-AAGGAGATCCGA 888

RESULT 9
BQ964179
LOCUS       BQ964179               988 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_10055799 NIH_MGC_134 Mus musculus cDNA clone
            IMAGE:6509783 5', mRNA sequence.
ACCESSION   BQ964179
VERSION     BQ964179.1  GI:22379657
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 988)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14077  row: j  column: 24
            High quality sequence stop: 702.
FEATURES             Location/Qualifiers
     source          1..988
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6509783"
                     /tissue_type="undifferentiated limb"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_134"
                     /note="Vector: pCMV-SPORT6.1; Site_1: EcoRV; Site_2: NotI;
                     Cloned unidirectionally. Primer: Oligo dT. Average insert
                     size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note:
                     this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:              9.59e-135      Length:          988
Score:                  1339.00        Matches:         248
Percent Similarity:     95.49%         Conservative:    6
Best Local Similarity:  93.23%         Mismatches:      11
Query Match:            67.66%         Indels:          2
DB:                     5              Gaps:            0

US-09-774-954-2 (1-365) x BQ964179 (1-988)

Qy            3 AlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGly 22
                ||||||||||||||    ||||||||||||||||||||||||||||||||||||||||||
Db          152 GCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGCTTTGGG 211

Qy           23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          212 AACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGCACCTTG 271

Qy           43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisVal 62
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          272 GCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTCCATGTG 331

Qy           63 SerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeu 82
                |||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db          332 TCCTACCAAAAGTACTTCAAACTGGAGCCTCTCCAAGCCTACCATCGGGTTGTCAGCCTG 391

Qy           83 GluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyr 102
                ||||||||||||||    ||||||||:::||||||||||||||||||||||||||||||
Db          392 GAGGACTTCATGGAAAATCTGGCACCCTCCCACTGGCCCCCTGAGAAGCGAGTGGCATAC 451

Qy          103 CysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGly 122
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          452 TGCTTTGAGGTGGCAGCCCAGCGAAGTCCTGATAAGAAGACATGTCCCATGAAGGAAGGA 511

Qy          123 AsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPhe 142
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          512 AATCCTTTTGGCCCATTCTGGGACCAGTTTCATGTGAGTTTCAATAAGTCAGAACTGTTC 571

Qy          143 ThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerProLys 162
                ||||||||||||||||||||||||||:::||||||||||:::|||||||||      |||
Db          572 ACAGGCATTTCCTTCAGCGCCTCCTACAAAGAACAATGGACCCAGAGATTTCCTGCAAAA 631

Qy          163 GluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGluHis 182
                ||||||||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db          632 GAGCATCCTGTGCTCGCACTGCCTGNGGCCCCAGCACAGTTCCCTGTCCTGGAGGAACAC 691

Qy          183 ArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAlaGln 202
                |||   ||||||||||||||||||||||||||||||||||||:::||||||||||||
Db          692 AGGGAGCTCCAGAAGTACATGGTGTGGTCAGATGAGATGGTGAGGACGGGAGAGGCCCTG 751
```

```
Qy        203 IleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTrp 222
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        752 ATCAGTGCCCACCTCGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCCGACTGG 811

Qy        223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
              ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db        812 AAGAATGCCTGTGCCATGCTGGAGGATGGAACTGCAGGGTCACACTTCATGGCTTCCCCT 871

Qy        243 Gln-CysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPr 262
                  ||||||||||||||||||||||||||||    ||||||   ||||||||||||||||||
Db        872 TCNNTGTGTGGGCTATAGCCGCAGCACAGCGACCCCTCTCAC-ATGACCATGTGCCTCCC 930

Qy        262 oAspLeuLysGluIle 267
              ||||||||||||||||
Db        931 TGACCTGAAGGAAATC 946

RESULT 10
BX345989
LOCUS       BX345989                841 bp    mRNA    linear   EST 23-APR-2004
DEFINITION  BX345989 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
            Homo sapiens cDNA clone CS0DJ008YF12 5-PRIME, mRNA sequence.
ACCESSION   BX345989
VERSION     BX345989.2  GI:46549776
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 841)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 5, 2003 this sequence version replaced gi:30377055.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            9980.f
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0AJ008DC06QP1&c=9980.f.
FEATURES             Location/Qualifiers
     source          1. .841
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ008YF12"
                     /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
                     /cell_line="JURKAT"
                     /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
                     10-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Alignment Scores:
Pred. No.:              9.14e-134      Length:         841
Score:                  1329.00        Matches:        242
Percent Similarity:     99.18%         Conservative:   0
Best Local Similarity:  99.18%         Mismatches:     2
Query Match:            67.16%         Indels:         1
DB:                     5              Gaps:           0

US-09-774-954-2 (1-365) x BX345989 (1-841)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        109 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 168

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        169 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 228

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 288

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        289 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 348

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        349 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 408

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              |||||||||||||||||||||||||||||||    ||||||||||||||||||||||||
Db        409 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAC-CCAGATAAGAAGACGTGCCCCATGAAG 467

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        468 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 527

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        528 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 587

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        588 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 647

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        648 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 707

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        708 GSCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 767

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        768 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 827

Qy        241 SerProGlnCys 244
              ||||||||||||
Db        828 TCTCCGCAGTGT 839

RESULT 11
BQ936634
LOCUS       BQ936634                915 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_10032510 NCI_CGAP_Mam2 Mus musculus cDNA clone
            IMAGE:6485896 5', mRNA sequence.
ACCESSION   BQ936634
VERSION     BQ936634.1  GI:22352017
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 915)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
```

```
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM14028  row: g  column: 17
           High quality sequence stop: 698.
FEATURES             Location/Qualifiers
     source          1. .915
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6485896"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam2"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
ORIGIN

Alignment Scores:
Pred. No.:                2.73e-132      Length:          915
Score:                    1316.00        Matches:         236
Percent Similarity:       96.43%         Conservative:    7
Best Local Similarity:    93.65%         Mismatches:      9
Query Match:              66.50%         Indels:          0
DB:                       5              Gaps:            0

US-09-774-954-2 (1-365) x BQ936634 (1-915)

Qy          3 AlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGly 22
              |||||||||||||||   |||||||||||||||||||||||||||||||||||||||||||
Db        156 GCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGCTTTGGG 215

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        216 AACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGCACCTTG 275

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisVal 62
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        276 GCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTCCATGTG 335

Qy         63 SerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeu 82
              |||||||||||||||||||||||||||||||||||||||||||||||||:::||||||
Db        336 TCCTACCAAAAGTACTTCAAACTGGAGCCTCTCCAAGCCTACCATCGGGTTGTCAGCCTG 395

Qy         83 GluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyr 102
              |||||||||||||||   |||||||||:::|||||||||||||||||||||||||||||
Db        396 GAGGACTTCATGGAAAATCTGGCACCCTCCCACTGGCCCCCTGAGAAGCGAGTGGCATAC 455

Qy        103 CysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGly 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        456 TGCTTTGAGGTGGCAGCCCAGCGAAGTCCTGATAAGAAGACATGTCCCATGAAGGAAGGA 515

Qy        123 AsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPhe 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        516 AATCCTTTTGGGCCATTCTGGGACCAGTTTCATGTGAGTTTCAATAAGTCAGAACTGTTC 575

Qy        143 ThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerProLys 162
              ||||||||||||||||||||||||||:::|||||||||:::|||||||||      |||
Db        576 ACAGGCATTTCCTTCAGCGCCTCCTACAAAGAACAATGGACCCAGAGATTTCCTGCAAAA 635

Qy        163 GluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGluHis 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        636 GAGCATCCTGTGCTCGCACTGCCTGGGGCCCCAGCACAGTTCCCTGTCCTGGAGGAACAC 695

Qy        183 ArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAlaGln 202
              |||   |||||||||||||||||||||||||||||||||||||:::|||||||||||||
Db        696 AGGGAGCTCCAGAAGTACATGGTGTGGTCAGATGAGATGGTGAGGACGGGAGAAGCCCTG 755

Qy        203 IleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTrp 222
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        756 ATCAGTGCCCACCTCGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCCGACTGG 815

Qy        223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
              ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||||
Db        816 AAGAATGCCTGTGGCATGCTGAAGGATGGAACTGCAGGGTCACACTTCATGGCCTCCCCT 875

Qy        243 GlnCysValGlyTyrSerArgSerThrAlaAlaPro 254
              ||||||||||||||:::||||||||:::   |||
Db        876 CAGTGTGTGGGCTATAACCGCAGCACAAGCGACCCC 911

RESULT 12
BX449172
LOCUS       BX449172                898 bp    mRNA    linear   EST 06-MAY-2004
DEFINITION  BX449172 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
            CS0DM014YK11 5-PRIME, mRNA sequence.
ACCESSION   BX449172
VERSION     BX449172.2  GI:47067286
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 22, 2003 this sequence version replaced gi:31030014.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 9980.f
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0AM014AF06QP1&c=9980.f.
FEATURES             Location/Qualifiers
     source          1. .898
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DM014YK11"
                     /tissue_type="FETAL LIVER"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL LIVER"
                     /note="Organ: liver; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

Alignment Scores:
Pred. No.:                5.38e-131      Length:          898
Score:                    1304.00        Matches:         244
Percent Similarity:       96.08%         Conservative:    1
Best Local Similarity:    95.69%         Mismatches:      8
Query Match:              65.89%         Indels:          3
DB:                       5              Gaps:            0

US-09-774-954-2 (1-365) x BX449172 (1-898)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 198

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 258
```

```
Qy          41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
               ||||||||||||||   ||||||||||||||||||||||||||||||||||||||||||
Db         259 ACCTTGGCTGTCCCTCHTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 318

Qy          61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         319 CATGTGTCCTACCAGAAGTACTTYAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 378

Qy          81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         379 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 438

Qy         101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
               ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db         439 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAC-CCAGATAAGAAGACGTGCCCCATGAAG 497

Qy         121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
               ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db         498 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTAAGTGTCAACAAGTCGGAG 557

Qy         141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         558 CTTTTTACAGGCATTTCCTTCAGTGCTTCGTACAGAGAACAATGGAGCCAGAGATTTTCT 617

Qy         161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         618 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 677

Qy         181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGlu-MetValLysThrGlyGl 200
               ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||::
Db         678 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAAATGGTGAAGACGGGAGA 737

Qy         200 uAlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySe 220
               :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         738 CGCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTC 797

Qy         220 rAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMet-A 240
               ||||||||||||||||||   ||||||||||||||||||||||||||||||   ||| |
Db         798 TGACTGGAAGAACGCCTGTCCAATGCTGAAGGACGGGACTGCAGGCTCGCACTTAATGGG 857

Qy         240 laSerProGlnCysValGlyTyrSerArgSerThrAlaAla 253
               ||   ||||||||||||||   ||||||||||   |||
Db         858 CTYCTCCGCAGTGTGTGGGCDACAGCCGCAGCACACSGGCC 898

RESULT 13
CO774833
LOCUS       CO774833                942 bp    mRNA    linear   EST 04-AUG-2004
DEFINITION  ILLUMIGEN_MCQ_51950 Katze_MNLV Macaca nemestrina cDNA clone
            IBIUW:29960 5' similar to Bases 7 to 852 highly similar to human
            POFUT1 (Hs.178292), mRNA sequence.
ACCESSION   CO774833
VERSION     CO774833.1  GI:50976096
KEYWORDS    EST.
SOURCE      Macaca nemestrina (pig-tailed macaque)
  ORGANISM  Macaca nemestrina
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 942)
  AUTHORS   Katze,M.G., Thomas,M., Korth,M., Iadonato,S.P. and Magness,C.L.
  TITLE     Large-scale Rhesus Macaque cDNA Sequencing
  JOURNAL   Unpublished (2003)
COMMENT     Contact: C. Magness
            Illumigen Biosciences Inc.
            2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
            Tel: 2063780400
            Fax: 2063780408
            Email: cmagness@illumigen.com
            Sequenced on 2004.07.21. 611 Q20 bases. Library Preparation: Prof.
            Michael Katze Lab at University of Washington DNA Sequencing:
            Illumigen Biosciences Inc. For further information, see
            http://www.macaque.org
            PCR PRimers
            FORWARD: CCCTCACTAAAGGGAACAAAA
            BACKWARD: CACTATAGGGCGAATTGGGTA
            Insert Length: 942   Std Error: 0.00
            Plate: CL000531  row: E  column: 09
            Seq primer: CCCTCACTAAAGGGAACAAAA
            POLYA=No.
FEATURES              Location/Qualifiers
     source           1. .942
                      /organism="Macaca nemestrina"
                      /mol_type="mRNA"
                      /strain="Indian"
                      /db_xref="taxon:9545"
                      /clone="IBIUW:29960"
                      /sex="male"
                      /lab_host="Electromax DH10B"
                      /clone_lib="Katze_MNLV"
                      /note="Organ: liver; Vector: pDONR 222; Site_1: BsrG I;
                      Site_2: BsrG I; Created from CloneMiner cDNA Library
                      Construction kit (catalog #18249-029)"
ORIGIN

Alignment Scores:
Pred. No.:              5.8e-131        Length:         942
Score:                  1304.00         Matches:        245
Percent Similarity:     94.36%          Conservative:   6
Best Local Similarity:  92.11%          Mismatches:     11
Query Match:            65.89%          Indels:         4
DB:                     7               Gaps:           0

US-09-774-954-2 (1-365) x CO774833 (1-942)

Qy           1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          95 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 154

Qy          21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         155 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGC 214

Qy          41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         215 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCCTTCACCAACCTC 274

Qy          61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         275 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 334

Qy          81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         335 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 394

Qy         101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         395 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 454

Qy         121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         455 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCTGAG 514

Qy         141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         515 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 574

Qy         161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         575 CCAAAGGAACATCCGGTGCTTGCCCTGCCGGGAGCCCCGGCCCAGTTCCCTGTCCTGGAA 634

Qy         181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
               ||||||||||||||||||||||||||||||||||||||:::|||||:::||||||||||
Db         635 GAACACAGACCACTGCAAAAGTACATGGTATGGTCAGACCAAATGGTGAGGACGGGAGAA 694

Qy         201 AlaGlnIleHisAlaHisLeuVal-ArgProTyrValGlyIleHisLeuArgIleGly-S 220
```

```
              ||||||      |||||||||||| ||||||||||||||||||    ||||||   ||| |
Db        695 GCCCAGAATCCCGCCCACCTTGTCCCGGCCCTATGTGGGCATTCCTCTACGTAATGGCTT 754

Qy        220 erAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        755 CTGACTGGAAAAATGCCTGTGCTATGCTAAAGGACGGAACTGGCAGCTCGACTTTCTTGG 814

Qy        240 laSerProGlnCysValGly-TyrSerArgSerThrAlaAla-ProLeuThrMetThrMe 259
              ||||||||:::|||   ||| |||:::   :::|||||||| ||||||      |||||
Db        815 CCTCTCCCGAATGTGGGGGCCTACGCCCCCAACACAGCCGCCCCCCCTCCCGAAGACTAT 874

Qy        259 tCysLeuPro 262
              ||||||||||
Db        875 GTGCCTGCCT 884

RESULT 14
AL135434
LOCUS       AL135434                719 bp    mRNA    linear   EST 04-SEP-2003
DEFINITION  DKFZp762O0115_r1 762 (synonym: hmel2) Homo sapiens cDNA clone
            DKFZp762O0115 5', mRNA sequence.
ACCESSION   AL135434
VERSION     AL135434.1  GI:6603621
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 719)
  AUTHORS   Ansorge,W., Wirkner,U., Mewes,W., Weil,B. and Wiemann,S.
  TITLE     EST (Ansorge,W., Wirkner,U., Mewes,H.W., Weil,B. and Wiemann,S.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone (DKFZp762O0115) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .719
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp762O0115"
                     /tissue_type="melanoma (MeWo cell line)"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="762 (synonym: hmel2)"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
ORIGIN

Alignment Scores:
Pred. No.:              1.16e-127      Length:         719
Score:                  1272.00        Matches:        239
Percent Similarity:     100.00%        Conservative:   0
Best Local Similarity:  100.00%        Mismatches:     0
Query Match:            64.27%         Indels:         0
DB:                     1              Gaps:           0

US-09-774-954-2 (1-365) x AL135434 (1-719)

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 61

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         62 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 121

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 181

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 241

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 301

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 361

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 421

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 481

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 541

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 601

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 661

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProPro 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCT 718

RESULT 15
BG747760
LOCUS       BG747760                808 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602705255F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842061 5',
            mRNA sequence.
ACCESSION   BG747760
VERSION     BG747760.1  GI:14058413
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 808)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1675  row: n  column: 14
            High quality sequence stop: 808.
FEATURES             Location/Qualifiers
     source          1. .808
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
```

```
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4842061"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_43"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
ORIGIN

Alignment Scores:
Pred. No.:              9.11e-124       Length:         808
Score:                  1237.00         Matches:        234
Percent Similarity:     96.71%          Conservative:   1
Best Local Similarity:  96.30%          Mismatches:     5
Query Match:            62.51%          Indels:         4
DB:                     4               Gaps:           0

US-09-774-954-2 (1-365) x BG747760 (1-808)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         82 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 141

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        142 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 201

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 261

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 321

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        322 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 381

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 441

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        442 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 501

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        502 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 561

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        562 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 621

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThr-GlyGl 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        622 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGGAGA 681

Qy        200 uAlaGlnIleHisAlaHisLeuValArgProTyrVal-GlyIleHisLeuArgIleGlyS 220
              ||||||||||||||   |||||||||   |||||| ||||||:::||||||||||||
Db        682 GGCCCAGATTCATGCCCAACTGGTCCGGACCTATGTTGGGCATCAATCTGCGCATTGGTC 741

Qy        220 erAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySer-HisPheMet 239
              ||||||||||||||||||||||||||||||||||      |||||| |||||||||
Db        742 T-GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGGACTGAGGCTCGGCACTTCATG 800
```

```
Qy        240 Ala 240
              |||
Db        801 GCT 803
```

Search completed: October 26, 2005, 15:17:12
Job time : 6398.13 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:21:43 ; Search time 45.7187 Seconds
(without alignments)
768.157 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....QGRPSSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	541.5	27.4	474	2	T15511	hypothetical prote
2	114	5.8	594	2	A86309	F20D23.3 protein -
3	105.5	5.3	383	2	T14795	hypothetical prote
4	101.5	5.1	732	1	S07624	acylaminoacyl-pept
5	96.5	4.9	432	2	T37509	hypothetical prote
6	94.5	4.8	1209	2	T31657	reverse transcript
7	93.5	4.7	458	2	G83690	hypothetical prote
8	91.5	4.6	1292	2	T31462	probable magnesium
9	90.5	4.6	445	2	I46225	mucin A, tracheal
10	90	4.5	437	2	F59099	hypothetical prote
11	89.5	4.5	218	4	S37459	self-incompatabili
12	89	4.5	650	2	T38692	probable serine/th
13	88.5	4.5	877	2	T03098	p97 protein - Toxo
14	88	4.4	380	2	T40454	branched-chain-ami
15	88	4.4	427	2	T04869	transforming prote
16	87.5	4.4	1246	2	G64899	nitrate reductase
17	87	4.4	407	2	D95211	tryptophan synthas
18	87	4.4	520	2	F91080	hypothetical prote
19	87	4.4	520	2	G85925	hypothetical prote
20	86.5	4.4	492	2	T29439	hypothetical prote
21	86.5	4.4	1189	2	S00944	parasporal crystal
22	86.5	4.4	1246	2	G90887	cryptic nitrate re
23	86.5	4.4	1246	2	B85730	cryptic nitrate re
24	85.5	4.3	328	2	S39599	class I histocompa
25	85.5	4.3	770	2	AI2421	hypothetical prote
26	85	4.3	421	2	T47393	hypothetical prote
27	85	4.3	732	1	JC4655	acylaminoacyl-pept
28	85	4.3	859	2	S69700	hypothetical prote
29	84.5	4.3	410	2	E69168	probable transamin
30	84.5	4.3	3519	2	S43048	polyketide synthas
31	84	4.2	1117	2	AE0075	hypothetical prote
32	83.5	4.2	283	2	S44848	K06H7.8 protein -
33	83.5	4.2	390	2	E86411	protein F1K23.18 [
34	83.5	4.2	518	2	JC7387	testis-abundant fi
35	83.5	4.2	1400	2	T33758	hypothetical prote
36	83	4.2	255	2	AH2479	transcription regu
37	83	4.2	349	1	E64096	UDPglucose-hexose-
38	83	4.2	554	2	S67452	hypothetical prote
39	83	4.2	803	2	AB2088	hypothetical prote
40	83	4.2	1242	2	AB0672	respiratory nitrat
41	82.5	4.2	263	2	JH0635	type II site-speci
42	82.5	4.2	534	2	F96713	unknown protein T6
43	82.5	4.2	659	2	AB1293	ABC transporter (p
44	82	4.1	520	2	T20226	hypothetical prote
45	82	4.1	623	2	B81037	lipopolysaccharide

ALIGNMENTS

RESULT 1
T15511
hypothetical protein C15C7.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15511
R;Leimbach, D.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid C15C7.
A;Reference number: Z18363
A;Accession: T15511
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-474 <LEI>
A;Cross-references: EMBL:U41528; NID:g1109795; PID:g1109800; PIDN:AAA83156.1; CESP:C15C7.
C;Genetics:
A;Gene: CESP:C15C7.1
A;Introns: 31/3; 67/2; 106/2; 166/1; 202/3; 235/3; 364/3; 410/3

Query Match 27.4%; Score 541.5; DB 2; Length 474;
Best Local Similarity 36.2%; Pred. No. 1.9e-38;
Matches 125; Conservative 60; Mismatches 107; Indels 53; Gaps 11;

```
Qy         19 GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHR 78
              |||||| | ||| ||||| |:||| :| :||::|   | | : : ::  |::  :  | |
Db        166 GRFGNQVDQFLGVLAFAKALDRTLVLPNFIEFKH---PETKM-IPFEFLFQVGTVAKYTR 221

Qy         79 VISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNK 138
              |::::| :|: |||:    | |  :: :|:       |  |||||||:|||  |||
Db        222 VVTMQEFTKKIMPTHFVGTPRQA-IYDKSAEPG-----CHSKEGNPFGPYWDQIDVSFVG 275

Qy        139 SELFTGI--SFSASY---REQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSD 193
              | |  |   |  :    |::| ::|  :|:||||   |||  ||   :   :|||: ||
Db        276 DEYFGDIPGGFDLNQMGSRKKWLEKFPSEEYPVLAFSSAPAPFPSKGKVWSIQKYLRWSS 335

Qy        194 EMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA 253
               : :  :  | |:| :|:| :|||  :||   |  : | |       || ||:|
Db        336 RITEQAKKFISANLAKPFVAVHLRNDADWVRVCEHI-DTTTNRPLFASEQCLGEGHHLGT 394

Qy        254 PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEV 313
              ||  :| |  ::|                              |:|     ::    :|:
Db        395 -LTKEICSPSKQQI---------------------------LEQ-----IEAHRQEPDD 420

Qy        314 AQVDLYILGQADHFIGNCVSSFTAFVKRERDLQG---RPSSFFGM 355
                | |:|:|| |:|||||:|:  ||||||  |   |||:|||:
Db        421 MYTSLAIMGRADLFVGNCVSTFSHIVKRERDHAGQSPRPSAFFGI 465
```

RESULT 2
A86309
F20D23.3 protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)

```
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: A86309
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A86309
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-594 <STO>
A;Cross-references: UNIPROT:Q9SHI5; GB:AE005172; NID:g5734743; PIDN:AAD50008.1; GSPDB:GN
C;Genetics:
A;Map position: 1

  Query Match             5.8%;  Score 114;  DB 2;  Length 594;
  Best Local Similarity   20.1%;  Pred. No. 0.1;
  Matches   84;  Conservative   50;  Mismatches  160;  Indels  124;  Gaps   19;

Qy          6 WDP-AGYLLYCPCMGRFGNQADHFL---GSLAFAKLLNRTLAVP-PWIEYQHHKPPFTNL 60
              | |      |: |:   |  ::| :     : || ||:| | :|    :||
Db        209 WKPRPDKFLFAICLS--GQMSNHLICLEKHMFFAALLDRVLVIPSSKFDYQ--------- 257

Qy         61 HVSYQKYFKLEPLQ---AYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTC 117
                | |   :| :        ||| : | |         :|  |:          |: |
Db        258 ---YDKVIDIERINTCLGRTVVISFDQFKE-------IDKKNNAHI---------DRFIC 298

Qy        118 PMKEGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQR------------------- 158
              :    |   : |: |:   |    |:|          ||:
Db        299 YVSSPQPC--YVDEDHIKKLKG---LGVSIGGKLEAPWSEDIKKPTKRTSQEVVEKFKSD 353

Qy        159 ----------FSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQIHAHLV 208
                       ::  |  ::  || |          | :  :| :   ::|         |
Db        354 DGVIAIGDVFYADMEQDLVMQPGGPINHKCKTLIEPSRLILVTAQRFIQT-------FLG 406

Qy        209 RPYVGIHLR---IGSDWKNACA----MLKDGTAGSHFMASPQ----CVGYSRSTAAPLTM 257
              : :: :|||        |    |    |  |  | :  :  |:    |   | |
Db        407 KNFISLHLRRHGFLKFWLQNLAIVAIMYLDSVAHTQLIGIPEHLSTCNAKSPS------- 459

Qy        258 TMCLPDLKEIQRAVKLWVRSLDAQSVYVATD-SESYVPELQQLFKGKVKVVSL---KPE- 312
              |   : :    :   |   :|  :|:|| :||    || |     ||| |    |:
Db        460 --CFYPIPQAADCISRMVERANAPVIYLSTDAAESETGLLQSLVVVDGKVVPLVKRPPQN 517

Qy        313 --------------------VAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPS 350
                                 | :|  |   :  |||   |:||  : | |   | |
Db        518 SAEKWDSLLYRHGIEDDSQVYAMLDKTICAMSSVFIGASGSTFTEDILRLRKDWGTSS 575


RESULT 3
T14795
hypothetical protein DKFZp434E171.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T14795
R;Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, August 1999
A;Reference number: Z18183
A;Accession: T14795
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-383 <POU>
A;Cross-references: EMBL:AL110285
A;Experimental source: adult testis; clone DKFZp434E171
C;Genetics:
A;Note: DKFZp434E171.1

  Query Match             5.3%;  Score 105.5;  DB 2;  Length 383;
  Best Local Similarity   21.1%;  Pred. No. 0.32;
  Matches   87;  Conservative   42;  Mismatches  126;  Indels  157;  Gaps   18;

Qy          3 AGSWDPAG-----------------------YLLY--CPCMGRFGNQADHFLGSLAFAK 36
              | || ||                       ||||   |  | |  : | ::   :  |
Db         16 AVSWPPASASGQEFWPGQSAADILSGAASRRRYLLYDVNPPEG-FNLRRDVYIRIASLLK 74

Qy         37 LLNRT----LAVPPWIEYQH-HKPPFTNLHVSYQKYFKLEPLQAYHRVISLEDFMEKLAP 91
                | :|    | :|||    |   |    : : : ::| | |     ||  | |   :|
Db         75 TLLKTEEWVLVLPPWGRLYHWQSPDIHQVRIPWSEFFDLPSLNKNIPVIEYEQF---IAE 131

Qy         92 THWPPEKRVAYCFEVAAQRSPDKKTCPMKEGN-----PFGPFWDQFHVSFNKSELFTG-- 144
              :  |   :| |  :  |:           |||          |  ||   | :| | : |
Db        132 SGGPFIDQV-YVLQSYAE--------GWKEGTWEEKVDERPCIDQLLYSQDKHEYYRGWF 182

Qy        145 ---------------ISFSASY----------------------------REQWSQRFS 160
                            :  |||                           :| |  | |
Db        183 WGYEETRGLNVSCLSVQGSASIVAPLLLRNTSARSVMLDRAENLLHDHYGGKEYWDTRRS 242

Qy        161 PKEHPVLALPGAPAQFPVLEEHR-------------PLQKYMVWSDEMVKTGEAQIHAHL 207
                        :  |     | :| |           | |:   |     || | |     |
Db        243 --------MVFARHLREVGDEFRSRHLNSTDDADRIPFQE--DWMKMKVKLGSA-----L 287

Qy        208 VRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLKEI 267
                ||:|:|||            ||   |                      :| |:
Db        288 GGPYLGVHLR-----------RKDFIWGHR--------------------QDVPSLEGA 315

Qy        268 QRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLY 319
              | ::  :::     |:||||:     | ::| |   ::|  :|   :::||
Db        316 VRKIRSLMKTHRLDKVFVATDAVR--KEYEELKKLLPEMVRFEPTWEELELY 365


RESULT 4
S07624
acylaminoacyl-peptidase (EC 3.4.19.1) - rat
N;Alternate names: acyl-peptide hydrolase; acylamino-acid-releasing enzyme
C;Species: Rattus norvegicus (Norway rat)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 09-Jul-2004
C;Accession: A33706; S07624
R;Kobayashi, K.; Lin, L.W.; Yeadon, J.E.; Klickstein, L.B.; Smith, J.A.
J. Biol. Chem. 264, 8892-8899, 1989
A;Title: Cloning and sequence analysis of a rat liver cDNA encoding acyl-peptide hydrolas
A;Reference number: A33706; MUID:89255359; PMID:2722805
A;Accession: A33706
A;Molecule type: mRNA
A;Residues: 1-732 <KOB>
A;Cross-references: UNIPROT:P13676; GB:J04733; NID:g202931; PIDN:AAA88506.1; PID:g202932
A;Note: parts of this sequence, including the amino end of the mature protein, were dete
R;Lin, L.W.; Lee, F.J.S.; Smith, J.A.
Nucleic Acids Res. 17, 4397-4400, 1989
A;Title: Structural organization of the rat acyl-peptide hydrolase gene.
A;Reference number: S07624; MUID:89296508; PMID:2578023
A;Accession: S07624
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-732 <LIN>
A;Cross-references: EMBL:X14915
C;Genetics:
A;Introns: 4/3; 49/1; 91/2; 122/3; 148/1; 202/3; 248/3; 279/2; 293/1; 333/3; 354/1; 386/
C;Superfamily: acylaminoacyl-peptidase
C;Keywords: blocked amino end; homotetramer; hydrolase; omega peptidase
F;1/Modified site: blocked amino end (Met) (probably acetylated) #status experimental
F;118,291,443/Modified site: lysine derivative (Lys) #status experimental
F;587,675,707/Active site: Ser, Asp, His #status predicted

  Query Match             5.1%;  Score 101.5;  DB 1;  Length 732;
  Best Local Similarity   22.9%;  Pred. No. 1.6;
  Matches   81;  Conservative   32;  Mismatches  101;  Indels  139;  Gaps   23;

Qy         30 GSLAFAKLLN----------RTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYH-- 77
              ||    |||           | ::|| ::    ||      ||:    : ||:   |
Db        405 GSAGSWKLLTIDKDLMVAQFSTPSLPPSLKVGFLPPPGKEQSVSWVSLEEAEPIPGIHWG 464
```

```
Qy          78 -RVISLEDFMEKLAPTHWPP-EKRVAYC---FE---VAAQRSPDKKTCP---MKEGNPFG 126
Db         465 VRVL-----------HPPPDQENVQYADLDFEAILLQPSNPPDKTQVPMVVMPHGGPHS 512

Qy         127 PF---WDQFHVSFNKSELFTGISFS---ASYREQWSQRFSPKEHPVLALPG-------AP 173
Db         513 SFVTAWMLFPAMLCK------MGFAVLLVNYR--GSTGFG--QDSILSLPGNVGHQDVKD 562

Qy         174 AQFPV----LEEHRPLQKYMVWSDEMVKTGEAQIHAHLV-------------RPYVGIHL 216
Db         563 VQFAVEQVLQEEHFDARRVAL----MGGSHGGFLSCHLIGQYPETYSACIARNPVINIAS 618

Qy         217 RIGS----DWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVK 272
Db         619 MMGSTDIPDW----CMVETG-----------FPYSNS---------CLPDL-------N 646

Qy         273 LWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILGQAD 325
Db         647 VWEEMLD--------------------KSPIKYI---PQVKTPVLLMLGQED 675


RESULT 5
T37509
hypothetical protein SPAC10F6.15 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T37509
R;Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, November 1997
A;Reference number: Z21719
A;Accession: T37509
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-432 <CON>
A;Cross-references: UNIPROT:O42654; EMBL:AL009197; PIDN:CAA15728.1; GSPDB:GN00066; SPDB:
A;Experimental source: strain 972h-; cosmid c10F6
C;Genetics:
A;Gene: SPDB:SPAC10F6.15
A;Map position: 1

  Query Match             4.9%;  Score 96.5;  DB 2;  Length 432;
  Best Local Similarity   21.3%;  Pred. No. 2.2;
  Matches   52;  Conservative   47;  Mismatches   80;  Indels   65;  Gaps   13;

Qy         163 EHPVLALPGAPAQFPVLEEHRPLQKYMVW---SDEMVKTGEAQIH-------------A 205
Db          45 ENPMAAITYADIEYP-HNTHFLLLDSISWIPCTCKILKLKEFELPDRFDIPNLFYDGWLS 103

Qy         206 HLVRPYVGIHLRIGSDWKNACAMLKD------------GTAGSHFMASPQCV-----GYS 248
Db         104 YVFHTFSGVYLKIGYEMDESGFCLPDQVYHLIDNHSCIQTAVSHLLRNHQALFKYLCDYL 163

Qy         249 RSTAAPLTMTMCLPDL-----KEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGK 303
Db         164 RSGEFPPLTVLIHHVMLYQYYPKSLQEA--LWA----AVEHYVNNSGEAY---------ST 208

Qy         304 VKVVSLKPEVAQVDLYILGQADHF-IG---NCV----SSFTAFVKRERDL--QGRPSSFF 353
Db         209 VQKLAVQKKIGNIRMYLVNPRDIFALGSTCNCIVVSSSNFQSYVQLRKPLLENNLPYEID 268

Qy         354 GMDR 357
Db         269 GFDK 272


RESULT 6
T31657
reverse transcriptase homolog - sea squirt (Ciona intestinalis)
C;Species: Ciona intestinalis
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C;Accession: T31657
R;Bird, A.P.; Clark, V.; Jones, S.J.; Leitgeb, S.; Dobson, R.; Tweedie, S.
submitted to the EMBL Data Library, December 1996
A;Reference number: Z21049
A;Accession: T31657
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1209 <BIR>
A;Cross-references: UNIPROT:P91581; EMBL:Z83760; NID:e1014349; PID:e289993; PIDN:CAB0604
C;Genetics:
A;Note: COS41.3

  Query Match             4.8%;  Score 94.5;  DB 2;  Length 1209;
  Best Local Similarity   20.0%;  Pred. No. 12;
  Matches   44;  Conservative   34;  Mismatches   83;  Indels   59;  Gaps    5;

Qy         150 SYREQWSQRFSPKEHPVLALPGAP-----------------AQFPVLEEHRPLQKYMVW 191
Db         679 AYYAKWISRFSHKIQPLLCIKKFPLTKDALVCFETLKTELAAASLGAIRSEEPLQVETDA 738

Qy         192 SDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRST 251
Db         739 SDYAI----AAVLSQSGRPVAYMSRTLNTCERNYPAIEKEATA----------------- 777

Qy         252 AAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKV---KVVS 308
Db         778 --------------VVEAVRKWSHFLKGKSFTLVTDQRSVSFMFDQRNRGKIKNSKILM 822

Qy         309 LKPEVAQVDLYILGQ--ADHFIGNCVSSFTAFVKRERDLQ 346
Db         823 WRLELSQFTYDIVHRPGRDNFVPDALSRVNALNRDLKELQ 862


RESULT 7
G83690
hypothetical protein BH0327 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
C;Accession: G83690
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiran
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: G83690
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-458 <STO>
A;Cross-references: UNIPROT:Q9KFY9; GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BAB040
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH0327

  Query Match             4.7%;  Score 93.5;  DB 2;  Length 458;
  Best Local Similarity   18.3%;  Pred. No. 4.3;
  Matches   51;  Conservative   50;  Mismatches   89;  Indels   89;  Gaps   11;

Qy          94 WPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELFTGISFSASYRE 153
Db         246 WPTLQ------QIFIKSSNSKLLSPLQA-------W-QF-VVFHYSQ-FRGLSNGKRGRS 289

Qy         154 QWSQ--------RFSPKEHPV----------LALPGAPAQFPVLEEHRPLQKYMVWSDEM 195
Db         290 HWPEPDSIRHMTRMAEEDHKTSVTIEKEDANIAFPRAQFGLPVIFQFHSKDRQDPFPTIL 349

Qy         196 VKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPL 255
Db         350 KPRGKDRLASRLI--------------------------------IKAVGVSERSAPPI 376

Qy         256 TMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQ 315
Db         377 VVTLNQPEIEEVELSLMKEVK----------TEEQKEVKEQLDKYKGKLKRDVVYPKNLN 426

Qy         316 VDLYILGQADHFIGNCVSSFTAFVKR-----ERDLQGRP 349
Db         427 YQKYPMKES-------TSAVEAFIKKGGSIWERALPEKP 458
```

RESULT 8
T31462
probable magnesium chelatase (EC 4.99.1.-) chain H BchH - Heliobacillus mobilis
C;Species: Heliobacillus mobilis
C;Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 09-Jul-2004
C;Accession: T31462
R;Xiong, J.; Inoue, K.; Bauer, C.E.
Proc. Natl. Acad. Sci. U.S.A. 95, 14851-14856, 1998
A;Title: Tracking molecular evolution of photosynthesis by characterization of a major p
A;Reference number: Z21036; MUID:99061957; PMID:9843979
A;Accession: T31462
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1292 <XIO>
A;Cross-references: UNIPROT:Q9ZGE5; EMBL:AF080002; NID:g3820536; PID:g3820560; PIDN:AAC8
C;Genetics:
A;Gene: bchH
C;Superfamily: Rhodobacter capsulatus magnesium-protoporphyrin O-methyltransferase
C;Keywords: lyase

Query Match 4.6%; Score 91.5; DB 2; Length 1292;
Best Local Similarity 20.3%; Pred. No. 24;
Matches 79; Conservative 54; Mismatches 118; Indels 139; Gaps 20;

```
Qy         39 NRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEK 98
Db        467 SREALIKAIIEDEESQLPSANLHVADR--------------ITVEQY-EKLTP-HW---R 507

Qy         99 RVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQR 158
Db        508 EIA---ETWGQPPGELNT----DGQDLLIFGRHF------GNVFIGVQPSFGY------- 547

Qy        159 FSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRI 218
Db        548 ---ESDPIKLLFTRSA-----SPHHGFMAFYRWLDKVYK---------------------- 578

Qy        219 GSDWKNACAMLKDGTAGS-HFMASPQCVG-------------------YS---------- 248
Db        579 ------ADALLHFGTHGALEFMPGHQ-VGLTDACWPDRLLGPVPNFYLYSVNNPSEATIA 631

Qy        249 -RSTAAPLTMTMCLP--------DLKEIQRAVKLW------------VRSLDAQSVYVAT 287
Db        632 KRRSAATLVSYLTPPAENAGLYKDLRDLKDLISMWGENRHNARGQSILETIIEKVLALHL 691

Qy        288 DSESYVP-ELQQLFKGK--VKVVSLKPEVAQVDLYILGQADH--FIGNCVSSFTAFVKRE 342
Db        692 DKDVPIPKEPGQEFIGKLYVYLTDLENRLIPTGLHILGEAPHAQTLGDYLMAISYFDRPE 751

Qy        343 RDLQGRP--------SSFFGMDRPPKLRDE 364
Db        752 LGVQSLAGLVSTALGSSFAEIERRSESGDQ 781
```

RESULT 9
I46225
mucin A, tracheal - dog (fragment)
C;Species: Canis lupus familiaris (dog)
C;Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 09-Jul-2004
C;Accession: PC1174; I46225
R;Shankar, V.; Tan, S.; Gilmore, M.S.; Sachdev, G.P.
Biochem. Biophys. Res. Commun. 189, 958-964, 1992
A;Title: Molecular cloning of the carboxy terminus of a canine tracheobronchial mucin.
A;Reference number: PC1174; MUID:93112073; PMID:1282002
A;Accession: PC1174
A;Molecule type: mRNA
A;Residues: 1-445 <SHA>
A;Cross-references: UNIPROT:Q01988; GB:L03387; NID:g164025; PIDN:AAA30875.1; PID:g164026
C;Keywords: glycoprotein
F;86,207,345/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 4.6%; Score 90.5; DB 2; Length 445;
Best Local Similarity 20.1%; Pred. No. 7.4;
Matches 58; Conservative 29; Mismatches 75; Indels 127; Gaps 14;

```
Qy        155 WSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEA-------QIHAHL 207
Db        171 WSRRRKPA--PVDNSPG-PSHWP----QRARRKHLFTLQTVNSNGTSDRSTFNEDTHA-- 221

Qy        208 VRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCL------ 261
Db        222 -QPYIAI------DWEPEMKKRYYDEVEAEGYVKHDCVGYVLK-KAPVRLQECIELFTTV 273

Qy        262 -----------PDLKEIQRAVK---LW-----------------------------VRS 277
Db        274 ETLEKENPWFCPTCKQHQLATKKLDLWMLPETLIIHLKRFSYTKFSREKLDTLVEFPIRD 333

Qy        278 LDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYIL---------------- 321
Db        334 LDFSEFVIKPQNES-----------------APELYKYDLIAVSNHYGGLRDGHYTTFA 375

Qy        322 -----GQADHFIGNCVS-----------SFTAFVKRE---RDLQGRPSS 351
Db        376 CNKDSGQSDYFDDNSVSPVTENQIESKAAYVLFYQRQDVARRLQPQPSS 424
```

RESULT 10
F59099
hypothetical protein pXO1-70 - Bacillus anthracis virulence plasmid pXO1
C;Species: Bacillus anthracis
C;Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 09-Jul-2004
C;Accession: F59099
R;Okinaka, R.T.; Cloud, K.; Hampton, O.; Hoffmaster, A.R.; Hill, K.K.; Keim, P.; Koehler,
J. Bacteriol. 181, 6509-6515, 1999
A;Title: Sequence and organization of pXO1, the large Bacillus anthracis plasmid harborir
A;Reference number: A59091; MUID:99445483; PMID:10515943
A;Accession: F59099
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-437 <OKI>
A;Cross-references: UNIPROT:Q9X340; GB:AF065404; NID:g4894216; PIDN:AAD32374.1; PID:g489
A;Experimental source: strain Sterne
C;Genetics:
A;Gene: pXO1-70
A;Genome: plasmid
C;Superfamily: Bacillus anthracis virulence plasmid pXO1 hypothetical protein pXO1-70

Query Match 4.5%; Score 90; DB 2; Length 437;
Best Local Similarity 20.7%; Pred. No. 8;
Matches 69; Conservative 35; Mismatches 136; Indels 94; Gaps 14;

```
Qy          2 PAGSWDPAGYLLYCPCMGRFGNQADHF------------------LGSLAFAKLLN--R 40
Db         17 PRGSWWEF-YRVTCPICHDTGNCMLHVSQEKVACTRVESKWIYGKNTGNPSYIHYINGKD 75

Qy         41 TLAVPPWIEYQ-HHKPPFTNLHVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKR 99
Db         76 KYQLPAADEVQIHDKKSNKELDVFNRKLMDFIPLQEHHHTHLLRD------------RKM 123

Qy        100 VAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELFTGI-SFSASYREQWSQR 158
Db        124 TEEQIQVRQYRSFLKQQIELEEDNTYTTVWEKLFNQIGNKNCWQGIPGFYEMKKGQLSLR 183

Qy        159 FSPKEHPVLALPGAPA-QFPVLEEHRPLQKYMVWSDEM-----VKTGEAQIHAHLVRPYV 212
Db        184 L---------MSGSPGILIPFRNQYNQIVGWQVRVDEVKNSVHVKSAPTGIQAELI---- 230

Qy        213 GIHLRIGSDWKNACAMLKDGTAGSHFMASPQCV-----GYSRSTAAPLTMTMCLPDLKEI 267
Db        231 --------EQPNVVKITKDG----------DCIFEGELEVSKKVEIP---------FQEG 263

Qy        268 QRAVK-------LWVRSLDAQSVYVATDSESYVP 294
Db        264 QIVVKIHKGQKYLWLSSANKNQGTGAGGSENPLP 297
```

```
RESULT 11
S37459
self-incompatability glycoprotein, mutant non-functional allele - Peruvian tomato (fragm
C;Species: Lycopersicon peruvianum (Peruvian tomato)
C;Date: 06-Jan-1995 #sequence_revision 31-Jan-1997 #text_change 09-Jul-2004
C;Accession: S37459
R;Royo, J.; Kunz, C.; Kowyama, Y.; Anderson, M.; Clarke, A.E.; Newbigin, E.
Proc. Natl. Acad. Sci. U.S.A. 91, 6511-6514, 1994
A;Title: Loss of a histidine residue at the active site of S-locus ribonuclease is assoc
A;Reference number: A58473; MUID:94294411; PMID:8022814
A;Accession: S37459
A;Molecule type: mRNA
A;Residues: 1-218 <ROY>
A;Cross-references: UNIPROT:Q40245; EMBL:Z26581; NID:g404333; PIDN:CAA81332.1; PID:g4043
A;Experimental source: strain LA 2157; style tissue
A;Note: submitted to the EMBL Data Library, September 1993
A;Note: part of this sequence, including the amino end of the mature protein, was determ
A;Note: the source is designated as Solanum peruvianum
C;Keywords: angiosperm reproduction; glycoprotein
F;1-17/Domain: signal sequence (fragment) #status predicted <SIG>
F;18-218/Product: self-incompatability glycoprotein #status predicted <MAT>
F;46,56/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             4.5%;  Score 89.5;  DB 4;  Length 218;
  Best Local Similarity   19.4%;  Pred. No. 3.6;
  Matches   45;  Conservative   43;  Mismatches  103;  Indels   41;  Gaps   10;

Qy         62 VSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKE 121
              |:  |||:  | ::|          :||      ||:        |       |:
Db          4 VFFMFLFALSPIYGYFELLEL--------VSTWP----ATYCYAYGCSRR------PIPN 45

Qy        122 GNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVL-- 179
                       |    |||:    :||   :::::   ||:   |         ::| |
Db         46 NFTINGLWPD-----NKSVILNNCNF-AKKEDRYTKITDPKKKSEL-----DKRWPQLRY 94

Qy        180 EEHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFM 239
              |:   ::|  :|  |:| |   |: :     |  : ::|     |:   :|  ||  :  :
Db         95 EKLYGIEKQDLWEKEFLKHGSCSINRYKQEAYFDLAMKI----KDRFDLL--GTLRNQGI 148

Qy        240 ASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSES 291
                         |  |::: :| || ||:     : :::   : :   | |:
Db        149 IPGSTYELDDIERAVKTVSIEVPSLKCIQKP----LGNVELNEIGICLDPEA 196

RESULT 12
T38692
probable serine/threonine-specific protein kinase (EC 2.7.1.-) SPAC3C7.06c [similarity]
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Aug-2004
C;Accession: T38692
R;Brown, D.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1997
A;Reference number: Z21805
A;Accession: T38692
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-650 <BRO>
A;Cross-references: UNIPROT:O14132; EMBL:Z99568; PIDN:CAB16737.1; GSPDB:GN00066; SPDB:SP
A;Experimental source: strain 972h-; cosmid c3C7
C;Genetics:
A;Gene: SPDB:SPAC3C7.06c
A;Map position: 1
C;Superfamily: protein kinase homology
C;Keywords: phosphotransferase

  Query Match             4.5%;  Score 89;  DB 2;  Length 650;
  Best Local Similarity   22.3%;  Pred. No. 16;
  Matches   71;  Conservative   50;  Mismatches  126;  Indels   72;  Gaps   16;

Qy          4 GSWDPAGYL-----LYCPCMG--RFGNQADHFL--GSLAFAKLLNRTLAVPPWIEYQHHK 54
              | || |           :  | |       ||    | |    :||||  :|:: |     |
Db        273 GIWDRAELLANKLGISLPKMAPLDFG---DLFSPPWNLAFASMLSQLLKWDP------AK 323

Qy         55 PPFTNLHVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDK 114
              |   : :  :            :  ||  |  :        :  ||:  :   ::  ||:
Db        324 RPTAEMCLDLE---------FCRVSAPADAVASKEEVNKNTDFRVSISYFPSSSSIPDE 373

Qy        115 KTCPMKEGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQ-----------RFSPK 162
                |  :|:  :     |   :|     ||   | |  |:  :|:  |         :|| :
Db        374 --CNTEEESRINPSTSKFLKQLNKG--FNG--FTKPFRKSRKQSKNRKNKSSVATQFSEE 427

Qy        163 EHPVLALPGAPAQFPVLEEHRP-------LQKYMVWSDEMVKTGEAQIHAHLVRPYVGIH 215
                  :      :   |||| : ||        |: ::: | |    : :|:       ||
Db        428 SEDIADSITSSTFFPVLPQIRPSTPLNLKLRNFIISSSEDSTSPKAK---EFDRPLPSTE 484

Qy        216 LRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWV 275
                :  : |:  |:| :                ||       |:| :   |: : || ||    | ::
Db        485 FLVAIN-KSQEALLNN---------SPN----SKSGSTQLSASTCLSDLISPQLSILSHE 530

Qy        276 RSLDAQSVYVATDSESYVP 294
                :  ||   | ::|   | |
Db        531 DKRENQS--VNSESSKYSP 547

RESULT 13
T03098
p97 protein - Toxoplasma gondii
C;Species: Toxoplasma gondii
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C;Accession: T03098
R;Matsuura, T.; Kasper, L.
Mol. Biochem. Parasitol. 90, 403-413, 1997
A;Title: Molecular analysis and characterization of a protein involved in the replicatio
A;Reference number: Z14838; MUID:98135655; PMID:9476788
A;Accession: T03098
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-877 <MAT>
A;Cross-references: UNIPROT:O15644; EMBL:AF005059; NID:g2581824; PIDN:AAC47857.1; PID:g25
A;Experimental source: strain RH
C;Function:
A;Description: involved in replication of intracellular Toxoplasma gondii
C;Superfamily: Toxoplasma gondii p97 protein

  Query Match             4.5%;  Score 88.5;  DB 2;  Length 877;
  Best Local Similarity   19.9%;  Pred. No. 27;
  Matches   74;  Conservative   44;  Mismatches  119;  Indels  135;  Gaps   20;

Qy          7 DPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQH-HKPPF-------- 57
              || |: ||   | | | :|  |  |     |     |    |||:: |   |
Db        392 DPEGWKLYAAHMQRGGRRA--FPGD--FDTWLRIHQAEAKWIEWRRVHSGPLNKELADPE 447

Qy         58 -------------TNLHVSYQKYFKLEPLQAYHRVIS-----------LEDFMEKLAPT 92
                           | :  :| :  ||   |  |  : |             |:|  ::     |
Db        448 GWELFRLRSAARGNTTVPATYAEWMKLPEQQVFVAVDAWAAWRQTHPGPLKDVLQADGAT 507

Qy         93 HWP-------------PEKRVAYC------FEVAAQRSPDKKTCPMKE--GNPFGPFWDQ 131
               |               |:   :         ::| |   ||   |::   :| |  |
Db        508 GWVLYRAMLRAADLDLPDDLATWLQDSGKHIKLAWQAWRDKHKGPLRRVLADPLG--W-- 563

Qy        132 FHVSFNKSELFTGISFSASYREQWSQRFSPKE-------------HPVLALP------- 170
                | |      :| |   :: |:|:|  :|                  | || |
Db        564 --VLFKNMVESSGASIEQTF-EEWAQSKAPSGNEWINWRRAHKGPLHEVLQDPEGWKLYR 620

Qy        171 ---------GAPAQF------PVLEEHRPLQKYMVWSDEMVKTGEAQIH-AHLV------ 208
                        |   |      |  ::        |:   |            ::|   ||
Db        621 ENLVRKNVVDVPQSFEDFLQQPDQQKAAATSKWSSW---------RRVHRGPLVTVLADK 671

Qy        209 ---RPYVGIHLRIG-------SDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMT 258
                  : ||  :  ::|        :| ::  | | | |   :   |        :  |     |||
Db        672 EGWQLYVNMMEKLGHKICSSFEEWSSSAQMHK-GRAAREWRA------WRRVHKGPLTTV 724

Qy        259 MCLPDLKEIQRA 270
               :  ||  :: :|
Db        725 LADPDGWDLYKA 736
```

```
RESULT 14
T40454
branched-chain-amino-acid transaminase (EC 2.6.1.42) precursor, mitochondrial - fission
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T40454; T40571
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Baker, S.; Mungall, K.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z21931
A;Accession: T40454
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-380 <LYN>
A;Cross-references: UNIPROT:O14370; EMBL:AL034382; PIDN:CAA22277.1; GSPDB:GN00067; SPDB:
A;Experimental source: strain 972h-; cosmid c428
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Davis, P.; Churcher, C.M.
submitted to the EMBL Data Library, July 1999
A;Reference number: Z21937
A;Accession: T40571
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-380 <WOO>
A;Cross-references: EMBL:AL096788; PIDN:CAB46675.1; GSPDB:GN00067; SPDB:SPBC582.12c
A;Experimental source: strain 972h-; cosmid c582
C;Genetics:
A;Gene: SPBC428.02c; SPBC582.12c
A;Map position: 2
A;Genome: nuclear
C;Superfamily: branched-chain-amino-acid transaminase BAT1
C;Keywords: aminotransferase; mitochondrion

  Query Match             4.4%;  Score 88;  DB 2;  Length 380;
  Best Local Similarity   19.6%;  Pred. No. 9.9;
  Matches   67;  Conservative   55;  Mismatches  121;  Indels   98;  Gaps   19;

Qy         52 HHKPPFTNLHVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRS 111
              |   |  : |:      :|:||  :  :   |:| : :    |  ||  :       :
Db          8 HGPKPMDSSHIKVTNVKELKPLPEWKSLKFGENFTDHMLIMKWNREKGWS-----TPEIV 62

Qy        112 PDKKTCPMKEGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPG 171
              | | |      | |    ||  |   | | |:    ::|::                 |
Db         63 PFGKLC-------FHPASSVFHYGF---ECFEGMK---AFRDE---------------KG 94

Qy        172 APAQFPVLEEHRPLQKYMVWSDEMVKTG---------EAQIHAHLVRPYV---------- 212
              | |      ||::     :: |: ||          |:: | ::| :|
Db         95 VPRLF------RPIKN----AERMLSTGTRISLPSFDPAEL-AEIIRKFVAHENRWVPDQ 143

Qy        213 ---GIHLR---IGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLKE 266
                 :::|   ||:|  |  :     |  : :|||    ||    |   : :|
Db        144 RGYSLYIRPTFIGTD--EALGVHHCDNAMLYVIASPVGPYYSSGFKA---VKLCCS---- 194

Qy        267 IQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILGQADH 326
              : :|: |        :     : | :|: :   ||  ::          |:: |  |:
Db        195 -EESVRAWPGGTGHYKL-GGNYAPSVLPQKEAAKKGYAQI----------LWLYGDEDY 241

Qy        327 F--IG--NCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRD 363
               :|  ||   || :: :  : :   :   ||  |   ||
Db        242 ITEVGTMNC---FTVWINKNGEKEIITAPLDGMILPGVTRD 279


RESULT 15
T04869
transforming protein myb homolog F28A21.180 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 16-Aug-2004
C;Accession: T04869
R;Bevan, M.; Mueller, M.W.; Muendlein, A.; Felber, R.; Bancroft, I.; Mewes, H.W.; Mayer,
submitted to the Protein Sequence Database, February 1999
A;Reference number: Z15387
A;Accession: T04869
A;Molecule type: DNA
A;Residues: 1-427 <BEV>
A;Cross-references: UNIPROT:Q9S7L2; EMBL:AL035526
A;Experimental source: cultivar Columbia; BAC clone F28A21
C;Genetics:
A;Map position: 4
A;Introns: 226/2; 268/3
A;Note: F28A21.180
C;Superfamily: myb DNA-binding repeat homology
F;212-263/Domain: myb DNA-binding repeat homology <MYB>

  Query Match             4.4%;  Score 88;  DB 2;  Length 427;
  Best Local Similarity   20.3%;  Pred. No. 11;
  Matches   71;  Conservative   53;  Mismatches  111;  Indels  114;  Gaps   19;

Qy         56 PFTNLHVSYQK----YFKLEP-------LQA-----------YHRVISLED--------- 84
              ||   | |||:    ::: :|        ||           :|: |:| |
Db         84 PFE--HCSYQENMVDFYETKPNLMNHHHFQAVENSYFTRNHHHHQEINLVDEHDDPMDLE 141

Qy         85 ----FMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSE 140
                  | :: |  :|| :     |:      ||: :|   : : :         ||||::
Db        142 QNNMMMMRMIPFDYPPTE----TFKPMNFVMPDEISCVSADNDCYRA------TSFNKTK 191

Qy        141 LF-----TGISFSASYRE---------QWS-----------QRFSPKE--HPVLALPGAP 173
               |     :  | |:|::|         ||:           :::  ::  |  |||
Db        192 PFLTRKLSSSSSSSSWKETKKSTLVKGQWTAEEDRVLIQLVEKYGLRKWSHIAQVLPGRI 251

Qy        174 AQFPVLEEH---RPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLK 230
              :       |   || |   ||:|             |  :  |  ||: |       |
Db        252 GKQCRERWHNHLRPDIKKETWSEEE-----------DRVLIEFHKEIGNKWAEIAKRLP 299

Qy        231 DGTAGS---HFMASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVAT 287
              |  |    |: |: :     |   :       | |          :::||:  ::
Db        300 GRTENSIKNHWNATKRRQFSKRKCRSKYPRPSLLQD----------YIKSLNMGALMA-- 347

Qy        288 DSESYVP----ELQQLFKGKVKVVSLKPEVAQVDLYILGQADHFIGNCV 332
              | ||      :   | |  ||::: :  : ::|  || |  :  ||
Db        348 ---SSVPARGRRRESNNKKKDVVVAVEEKKKEEEVY--GQ-DRIVPECV 390


Search completed: October 25, 2005, 15:36:41
Job time : 59.7187 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:25:36 ; Search time 59.9589 Seconds
(without alignments)
454.426 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....QGRPSSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1979	100.0	365	3	US-08-978-741-2	Sequence 2, Appli
2	1979	100.0	365	3	US-09-333-729A-3	Sequence 3, Appli
3	1979	100.0	397	3	US-08-978-741-6	Sequence 6, Appli
4	1979	100.0	397	3	US-09-333-729A-7	Sequence 7, Appli
5	1844	93.2	343	3	US-08-978-741-17	Sequence 17, Appl
6	1844	93.2	343	3	US-09-333-729A-13	Sequence 13, Appl
7	837	42.3	399	4	US-09-270-767-45921	Sequence 45921, A
8	541.5	27.4	474	3	US-08-978-741-8	Sequence 8, Appli
9	541.5	27.4	474	3	US-09-333-729A-12	Sequence 12, Appl
10	353	17.8	61	3	US-08-978-741-9	Sequence 9, Appli
11	353	17.8	61	3	US-09-333-729A-4	Sequence 4, Appli
12	343	17.3	200	4	US-09-270-767-61459	Sequence 61459, A
13	295	14.9	61	3	US-08-978-741-3	Sequence 3, Appli
14	295	14.9	61	3	US-09-333-729A-5	Sequence 5, Appli
15	101	5.1	28	3	US-08-978-741-14	Sequence 14, Appl
16	101	5.1	28	3	US-09-333-729A-14	Sequence 14, Appl
17	92.5	4.7	1271	4	US-09-540-236-3740	Sequence 3740, Ap
18	92.5	4.7	2074	4	US-09-491-356C-9	Sequence 9, Appli
19	87.5	4.4	344	3	US-09-298-886-11	Sequence 11, Appl
20	87.5	4.4	344	4	US-09-999-672-11	Sequence 11, Appl
21	87	4.4	555	4	US-09-248-796A-20471	Sequence 20471, A
22	86.5	4.4	771	2	US-08-742-753-2	Sequence 2, Appli
23	86.5	4.4	1189	1	US-08-602-737-2	Sequence 2, Appli
24	86.5	4.4	1189	2	US-08-980-071-2	Sequence 2, Appli
25	86.5	4.4	1189	2	US-08-980-071-4	Sequence 4, Appli
26	86.5	4.4	1189	2	US-08-980-071-6	Sequence 6, Appli
27	86.5	4.4	1189	2	US-08-980-071-8	Sequence 8, Appli
28	86.5	4.4	1189	2	US-08-980-071-10	Sequence 10, Appl
29	86.5	4.4	1189	2	US-08-980-071-12	Sequence 12, Appl
30	86.5	4.4	1189	2	US-08-980-071-59	Sequence 59, Appl
31	86.5	4.4	1189	2	US-08-980-071-61	Sequence 61, Appl
32	86.5	4.4	1189	2	US-08-757-536-2	Sequence 2, Appli
33	86.5	4.4	1189	2	US-08-757-536-4	Sequence 4, Appli
34	86.5	4.4	1189	2	US-08-757-536-6	Sequence 6, Appli
35	86.5	4.4	1189	2	US-08-757-536-8	Sequence 8, Appli
36	86.5	4.4	1189	2	US-08-757-536-10	Sequence 10, Appl
37	86.5	4.4	1189	2	US-08-757-536-12	Sequence 12, Appl
38	86.5	4.4	1189	3	US-09-314-093-2	Sequence 2, Appli
39	86.5	4.4	1189	3	US-09-314-093-4	Sequence 4, Appli
40	86.5	4.4	1189	3	US-09-314-093-6	Sequence 6, Appli
41	86.5	4.4	1189	3	US-09-314-093-8	Sequence 8, Appli
42	86.5	4.4	1189	3	US-09-314-093-10	Sequence 10, Appl
43	86.5	4.4	1189	3	US-09-314-093-12	Sequence 12, Appl
44	86.5	4.4	1189	3	US-09-314-093-59	Sequence 59, Appl
45	86.5	4.4	1189	3	US-09-314-093-61	Sequence 61, Appl

ALIGNMENTS

RESULT 1
US-08-978-741-2
; Sequence 2, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 365 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-978-741-2

Query Match 100.0%; Score 1979; DB 3; Length 365;
Best Local Similarity 100.0%; Pred. No. 2.4e-214;
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360

Qy        361 LRDEF 365
              |||||
Db        361 LRDEF 365


RESULT 2
US-09-333-729A-3
; Sequence 3, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 3
;   LENGTH: 365
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-333-729A-3

  Query Match           100.0%;  Score 1979;  DB 3;  Length 365;
  Best Local Similarity  100.0%;  Pred. No. 2.4e-214;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360

Qy        361 LRDEF 365
              |||||
Db        361 LRDEF 365


RESULT 3
US-08-978-741-6
; Sequence 6, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
;   APPLICANT:  Yang Wang, Michael W. Spellman
;   TITLE OF INVENTION:  O-Fucosyltransferase
;   NUMBER OF SEQUENCES:  17
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Genentech, Inc.
;     STREET:  1 DNA Way
;     CITY:  South San Francisco
;     STATE:  California
;     COUNTRY:  USA
;     ZIP:  94080
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  WinPatin (Genentech)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/978,741
;     FILING DATE:  26-No. 6100076-1997
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/792498
;     FILING DATE:  31
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Svoboda, Craig G.
;     REGISTRATION NUMBER:  39,044
;     REFERENCE/DOCKET NUMBER:  P1041P1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  650/225-1489
;     TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  397 amino acids
;     TYPE:  Amino Acid
;     TOPOLOGY:  Linear
US-08-978-741-6

  Query Match           100.0%;  Score 1979;  DB 3;  Length 397;
  Best Local Similarity  100.0%;  Pred. No. 2.8e-214;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         33 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 92

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 152

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 212

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 272

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 332

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 392
```

```
Qy        361 LRDEF 365
              |||||
Db        393 LRDEF 397


RESULT 4
US-09-333-729A-7
; Sequence 7, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 7
;   LENGTH: 397
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid insert encoded protein.
US-09-333-729A-7

  Query Match             100.0%;  Score 1979;  DB 3;  Length 397;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-214;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         33 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 92

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 152

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 212

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 272

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 332

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 392

Qy        361 LRDEF 365
              |||||
Db        393 LRDEF 397


RESULT 5
US-08-978-741-17
; Sequence 17, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  343 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-17

  Query Match             93.2%;  Score 1844;  DB 3;  Length 343;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-199;
  Matches  343;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         23 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRVISL 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRVISL 60

Qy         83 EDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELF 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELF 120

Qy        143 TGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQ 202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQ 180

Qy        203 IHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLP 262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLP 240

Qy        263 DLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILG 322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 DLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILG 300

Qy        323 QADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 365
              |||||||||||||||||||||||||||||||||||||||||||
Db        301 QADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 343


RESULT 6
US-09-333-729A-13
; Sequence 13, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 13
;   LENGTH: 343
```

```
;   TYPE: PRT
; · ORGANISM: Homo Sapien
US-09-333-729A-13

  Query Match             93.2%;  Score 1844;  DB 3;  Length 343;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-199;
  Matches  343;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         23 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRVISL 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRVISL 60

Qy         83 EDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELF 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELF 120

Qy        143 TGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQ 202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQ 180

Qy        203 IHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLP 262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLP 240

Qy        263 DLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILG 322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 DLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILG 300

Qy        323 QADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 365
              |||||||||||||||||||||||||||||||||||||||||||
Db        301 QADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 343


RESULT 7
US-09-270-767-45921
; Sequence 45921, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45921
;   LENGTH: 399
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-270-767-45921

  Query Match             42.3%;  Score 837;  DB 4;  Length 399;
  Best Local Similarity   44.9%;  Pred. No. 1.9e-85;
  Matches  168;  Conservative   62;  Mismatches  118;  Indels   26;  Gaps    9;

Qy          7 DPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQK 66
              || ||| |||||||||||||||||||||| ||||| :|||:||:  :   :  | :
Db         25 DPNGYLTYCPCMGRFGNQADHFLGSLAFAKALNRTLILPPWVEYR--RGELRSRQVPFNT 82

Qy         67 YFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFE-----VAAQRSPDKKTCPMKE 121
              ||::|||: |||||:: ||| ||   ||  :||::|::        :  |||  |  |:
Db         83 YFEVEPLKEYHRVITMADFMWHLADDIWPESERVSFCYKERYSLQQEKNDPDKPNCHAKD 142

Qy        122 GNPFGPPWDQFHVSFNKSELFTGISFSASYREQ---WSQRFSPKEHPVLALPGAPAQFPV 178
              ||||||||| ||: | :|| :  : |   :  :   |  ::  : :||||  |||| |||
Db        143 GNPFGPFWDTFHIDFVRSEFYAPLHFDVHHSNEAAKWQTKYPAESYPVLAFTGAPASFPV 202

Qy        179 LEEHRPLQKYMVWSDEMVKTGEAQIHAHLVR-PYVGIHLRIGSDWKNACAMLKDGTAGSH 237
                |:  ||:|: ||    :  :  |   | |  ::||||| | ||  ||  :||     |
Db        203 QLENCKLQRYLQWSQRYREASKDFIREQLPRGAFLGIHLRNGIDWVRACEHVKD---SQH 259

Qy        238 FMASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSL-------DAQSVYVATDSE 290
              |||||:||     | |   :|:|  : | | :|  ::::       : :||:||:||
```

```
Db        260 LFASPQCLGYKNERGA-LYPELCMPSKEAIIRQLKRTIKNVRQTQPDNEIKSVFVASDSN 318

Qy        291 SYVPELQ-QLFKGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRP 349
                : ||   | :  : |  |  :   :|| ||||::||||||:||::|||||||: | |
Db        319 HMIGELNTALSRMGISVHKLTEDDPYLDLAILGQSNHFIGNCISSYSAFVKRERDVHGFP 378

Qy        350 SSFFGMDRPPKLRD 363
              | |:|    || :|
Db        379 SYFWGF---PKEKD 389


RESULT 8
US-08-978-741-8
; Sequence 8, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  474 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-8

  Query Match             27.4%;  Score 541.5;  DB 3;  Length 474;
  Best Local Similarity   36.2%;  Pred. No. 5.6e-52;
  Matches  125;  Conservative   60;  Mismatches  107;  Indels   53;  Gaps   11;

Qy         19 GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHR 78
              |||||| | ||| ||||| |:||| :| :||::|   | | : : ::  |::  :  | |
Db        166 GRFGNQVDQFLGVLAFAKALDRTLVLPNFIEFKH---PETKM-IPFEFLFQVGTVAKYTR 221

Qy         79 VISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNK 138
              |:::::| :|: |||:    | |  :: :|:       |  |||||||||:|||  |||
Db        222 VVTMQEFTKKIMPTHFVGTPRQA-IYDKSAEPG-----CHSKEGNPFGPYWDQIDVSFVG 275

Qy        139 SELFTGI--SFSASY---REQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSD 193
              | |  |   |  :    |::| ::|  :|:||||   ||| ||   :   :|||: ||
Db        276 DEYFGDIPGGFDLNQMGSRKKWLEKFPSEEYPVLAFSSAPAPFPSKGKVWSIQKYLRWSS 335

Qy        194 EMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA 253
              : :  :  | |:| :|:| :|||  :||    |  : | |       || ||:|
Db        336 RITEQAKKFISANLAKPFVAVHLRNDADWVRVCEHI-DTTTNRPLFASEQCLGEGHHLGT 394

Qy        254 PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEV 313
```

```
              ||  :| |  ::|                              |:|     ::    :|:
Db        395 -LTKEICSPSKQQI----------------------------LEQ-----IEAHRQEPDD 420

Qy        314 AQVDLYILGQADHFIGNCVSSFTAFVKRERDLQG---RPSSFFGM 355
                | |:|:|| |:|||||:|:  ||||||  |   |||:|||:
Db        421 MYTSLAIMGRADLFVGNCVSTFSHIVKRERDHAGQSPRPSAFFGI 465


RESULT 9
US-09-333-729A-12
; Sequence 12, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 12
;   LENGTH: 474
;   TYPE: PRT
;   ORGANISM: Caenorhabditis Elegans
US-09-333-729A-12

  Query Match             27.4%;  Score 541.5;  DB 3;  Length 474;
  Best Local Similarity   36.2%;  Pred. No. 5.6e-52;
  Matches  125;  Conservative   60;  Mismatches  107;  Indels   53;  Gaps   11;

Qy         19 GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHR 78
              |||||| | ||| ||||| |:||| :| :||::|   | | : : ::  |::  :  | |
Db        166 GRFGNQVDQFLGVLAFAKALDRTLVLPNFIEFKH---PETKM-IPFEFLFQVGTVAKYTR 221

Qy         79 VISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNK 138
              |:::::| :|: |||:    | |  :: :|:       |  ||||||||:|||  |||
Db        222 VVTMQEFTKKIMPTHFVGTPRQA-IYDKSAEPG-----CHSKEGNPFGPYWDQIDVSFVG 275

Qy        139 SELFTGI--SFSASY---REQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSD 193
              | |  |   |  :    |::| ::|  :|:||||   ||| ||   :   :|||: ||
Db        276 DEYFGDIPGGFDLNQMGSRKKWLEKFPSEEYPVLAFSSAPAPFPSKGKVWSIQKYLRWSS 335

Qy        194 EMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA 253
               : :  :  | |:| :|:| :|||  :||    |  : | |       || ||:|
Db        336 RITEQAKKFISANLAKPFVAVHLRNDADWVRVCEHI-DTTTNRPLFASEQCLGEGHHLGT 394

Qy        254 PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEV 313
              ||  :| |  ::|                          |:|     ::    :|:
Db        395 -LTKEICSPSKQQI----------------------------LEQ-----IEAHRQEPDD 420

Qy        314 AQVDLYILGQADHFIGNCVSSFTAFVKRERDLQG---RPSSFFGM 355
                | |:|:|| |:|||||:|:  ||||||  |   |||:|||:
Db        421 MYTSLAIMGRADLFVGNCVSTFSHIVKRERDHAGQSPRPSAFFGI 465
```

1 - 61 of 2

```
RESULT 10
US-08-978-741-9
; Sequence 9, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  61 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-9

  Query Match             17.8%;  Score 353;  DB 3;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 4e-32;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 11
US-09-333-729A-4
; Sequence 4, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 4
;   LENGTH: 61
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-333-729A-4

  Query Match             17.8%;  Score 353;  DB 3;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 4e-32;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 12
US-09-270-767-61459
```

```
; Sequence 61459, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61459
;   LENGTH: 200
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-270-767-61459

  Query Match             17.3%;  Score 343;  DB 4;  Length 200;
  Best Local Similarity   40.6%;  Pred. No. 3.5e-30;
  Matches   80;  Conservative   34;  Mismatches   67;  Indels   16;  Gaps    6;

Qy        176 FPVLEEHRPLQKYMVWSDEMVKTGEAQIHAHLVR-PYVGIHLRIGSDWKNACAMLKDGTA 234
              |||  |:  ||:|: ||    :  :   |   | |  ::||||| | ||  ||  :||
Db          1 FPVQLENCKLQRYLQWSQRYREASKDFIREQLPRGAFLGIHLRNGIDWVRACEHVKD--- 57

Qy        235 GSHFMASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSL-------DAQSVYVAT 287
                |  |||||:||     | |   :|:|  : | | :|  ::::       : :||:||:
Db         58 SQHLFASPQCLGYKNERGA-LYPELCMPSKEAIIRQLKRTIKNVRQTQPDNEIKSVFVAS 116

Qy        288 DSESYVPELQ-QLFKGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQ 346
              ||   : ||   | :  : |  |  :   :|| |||::|||||:||::||||||||:
Db        117 DSNHMIGELNTALSRMGISVHKLTEDDPYLDLAILGQSNHFIGNCISSYSAFVKRERDVH 176

Qy        347 GRPSSFFGMDRPPKLRD 363
              | || |:|    || :|
Db        177 GFPSYFWGF---PKEKD 190


RESULT 13
US-08-978-741-3
; Sequence 3, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  61 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-3

  Query Match             14.9%;  Score 295;  DB 3;  Length 61;
  Best Local Similarity   91.5%;  Pred. No. 1.4e-25;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||
Db          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 14
US-09-333-729A-5
; Sequence 5, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 5
;   LENGTH: 61
;   TYPE: PRT
;   ORGANISM: Cricetulus Grieseus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 15, 17, 38
;   OTHER INFORMATION: unknown amino acid
US-09-333-729A-5

  Query Match             14.9%;  Score 295;  DB 3;  Length 61;
  Best Local Similarity   91.5%;  Pred. No. 1.4e-25;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||
Db          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 15
US-08-978-741-14
; Sequence 14, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
```

; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-978-741-14

Query Match 5.1%; Score 101; DB 3; Length 28;
Best Local Similarity 90.0%; Pred. No. 0.00031;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 MPAGSWDPAGYLLYCPCMGR 20
              |||||||||||||| | |||
Db          9 MPAGSWDPAGYLLYXPXMGR 28
```

Search completed: October 25, 2005, 15:42:47
Job time : 60.9589 secs

Wallka, M.
09/774,954
Seq. IDs 2, 3 + 9

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:20:23 ; Search time 232.341 Seconds
(without alignments)
607.588 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....QGRPSSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1979	100.0	365	2	AAW80571	Aaw80571 Human hea
2	1979	100.0	388	8	ABM82223	Abm82223 Tumour-as
3	1979	100.0	397	2	AAW80573	Aaw80573 Human O-f
4	1867.5	94.4	417	4	AAM78377	Aam78377 Human pro
5	1855.5	93.8	417	4	AAM79361	Aam79361 Human pro
6	1844	93.2	343	2	AAW80577	Aaw80577 Partial h
7	831	42.0	402	4	ABB63790	Abb63790 Drosophil
8	809	40.9	165	4	AAE03110	Aae03110 Human gen
9	809	40.9	165	5	ABG63619	Abg63619 Human alb
10	809	40.9	165	8	ADL76884	Adl76884 Albumin f
11	353	17.8	61	2	AAW80574	Aaw80574 N-termina
12	295	14.9	61	2	AAW80572	Aaw80572 N-termina
13	164.5	8.3	328	5	ABG95669	Abg95669 Human nuc
14	159.5	8.1	490	4	ABB69800	Abb69800 Drosophil
15	159.5	8.1	490	8	ADS96504	Ads96504 Drosophil
16	142	7.2	429	5	ABG95651	Abg95651 Human nuc
17	120	6.1	150	3	AAY87325	Aay87325 Human sig
18	105.5	5.3	428	3	AAB40332	Aab40332 Human ORF
19	104	5.3	566	3	AAG55221	Aag55221 Arabidops
20	102.5	5.2	260	6	ABJ25654	Abj25654 Aspergill
21	102	5.2	566	3	AAG60442	Aag60442 Arabidops
22	101.5	5.1	783	1	AAP94265	Aap94265 Sequence
23	101	5.1	28	2	AAW80575	Aaw80575 N-termina
24	100.5	5.1	772	8	ABM84900	Abm84900 Human dia
25	100.5	5.1	772	8	ABM84902	Abm84902 Human dia
26	100.5	5.1	772	8	ABM84903	Abm84903 Human dia
27	100.5	5.1	772	8	ABM84899	Abm84899 Human dia
28	100	5.1	207	7	ADE95590	Ade95590 Human NOV
29	100	5.1	207	7	ADE95588	Ade95588 Human NOV
30	98.5	5.0	1216	3	AAB42097	Aab42097 Human ORF
31	94.5	4.8	349	6	ADA21127	Ada21127 Human sec
32	94.5	4.8	764	8	ABM84901	Abm84901 Human dia
33	94.5	4.8	4773	7	ADJ95094	Adj95094 Novel NOV
34	94	4.7	1449	7	ADC71275	Adc71275 Human col
35	93	4.7	692	3	AAB49313	Aab49313 Mouse bro
36	92.5	4.7	1251	6	ABU35306	Abu35306 Protein e
37	92.5	4.7	1271	8	ADL06054	Adl06054 M. catarr
38	92.5	4.7	2074	3	AAY54319	Aay54319 Amino aci
39	92	4.6	348	6	ABU25276	Abu25276 Protein e
40	92	4.6	650	6	ABU45346	Abu45346 Protein e
41	91.5	4.6	344	7	ADB97786	Adb97786 Human alp
42	91	4.6	1381	5	ABG93234	Abg93234 C. albica
43	91	4.6	1381	5	ABP73947	Abp73947 Candida a
44	90.5	4.6	353	7	ADE95586	Ade95586 Human NOV
45	90.5	4.6	455	4	ABG25040	Abg25040 Novel hum

ALIGNMENTS

RESULT 1
AAW80571
ID AAW80571 standard; protein; 365 AA.
XX
AC AAW80571;
XX
DT 16-DEC-1998 (first entry)
XX
DE Human heart O-fucosyltransferase.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1. .61
FT /note= "N-terminal sequence claimed for in claim 4"
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
DR N-PSDB; AAV65632.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Example 1; Fig 12A; 90pp; English.
XX
CC This represents a human heart O-fucosyltransferase that can glycosylate
CC an epidermal growth factor (EGF) domain of a polypeptide with an
CC activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.
CC mutants with increased affinity for the EFG domains, are used in
CC diagnosis and treatment of conditions associated with overexpression of O
CC -fucosyltransferase, to promote survival of sensory (retinal) neurons.
CC Probes based on EGF domain polypeptide are used to detect gene

```
CC   amplification and expression. The expression can also be determined at
CC   the protein level using antibodies specific for O-fucosyltransferase
XX
SQ   Sequence 365 AA;

  Query Match             100.0%;  Score 1979;  DB 2;  Length 365;
  Best Local Similarity   100.0%;  Pred. No. 2.5e-205;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360

Qy        361 LRDEF 365
              |||||
Db        361 LRDEF 365


RESULT 2
ABM82223
ID   ABM82223 standard; protein; 388 AA.
XX
AC   ABM82223;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Tumour-associated antigenic target (TAT) polypeptide PRO83146, SEQ:5726.
XX
KW   Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW   tumour; diagnosis; cell proliferative disorder; breast cancer;
KW   colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW   central nervous system cancer; bladder cancer; pancreatic cancer;
KW   cervical cancer; melanoma; leukaemia; hybridisation probe;
KW   chromosome identification; chromosome mapping; gene mapping;
KW   gene therapy; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004030615-A2.
XX
PD   15-APR-2004.
XX
PF   29-SEP-2003; 2003WO-US028547.
XX
PR   02-OCT-2002; 2002US-0414971P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wu TD,  Zhang Z,  Zhou Y;
XX
DR   WPI; 2004-347921/32.
DR   N-PSDB; ACN40746.
XX
PT   New tumor-associated antigenic target polypeptides and nucleic acids,
PT   useful in preparing a medicament for treating or detecting a
PT   proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT   prostate cancer or tumor.
XX
PS   Claim 12; SEQ ID NO 5726; 7273pp; English.
XX
CC   The invention relates to human tumour-associated antigenic target (TAT)
CC   polypeptides, and their related nucleic acids. The TAT polypeptides are
CC   overexpressed in cancer tissues compared to normal tissues, and may thus
CC   serve as effective targets for the diagnosis and treatment of cancer in
CC   mammals. The invention also relates to nucleic acid and polypeptide
CC   sequences at least 80% identical to the TAT nucleic acids and
CC   polypeptides; expression vectors and host cells comprising a TAT nucleic
CC   acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC   molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC   TAT polypeptide; and methods and compositions for the treatment or
CC   diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC   antibodies, antagonists, binding molecules and compositions are useful
CC   for diagnosing or treating a cell proliferative disorder associated with
CC   increased TAT expression, particularly cancers such as breast cancer,
CC   colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC   cancer, pancreatic cancer, cervical cancer, cancers of the central
CC   nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC   used as hybridisation probes, in chromosome and gene mapping, in
CC   chromosome identification and in gene therapy. The present sequence
CC   represents a TAT polypeptide of the invention
XX
SQ   Sequence 388 AA;

  Query Match             100.0%;  Score 1979;  DB 8;  Length 388;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-205;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 83

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         84 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 143

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 203

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 263

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 323

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 383

Qy        361 LRDEF 365
              |||||
Db        384 LRDEF 388


RESULT 3
AAW80573
ID   AAW80573 standard; protein; 397 AA.
XX
AC   AAW80573;
XX
DT   16-DEC-1998  (first entry)
XX
DE   Human O-fucosyltransferase sequence expressed by a plasmid insertion.
XX
```

```
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         27. .32
FT                   /note= "polyhistidine tag"
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
DR   N-PSDB; AAV65633.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example; Fig 13B; 90pp; English.
XX
CC   This represents the amino acid sequence of the human heart O-
CC   fucosyltransferase expressed by a plasmid insertion. The human O-
CC   fucosyltransferase can glycosylate an epidermal growth factor (EGF)
CC   domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC   O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 397 AA;

  Query Match             100.0%;  Score 1979;  DB 2;  Length 397;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-205;
  Matches  365;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         33 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 92

Qy         61 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 HVSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMK 152

Qy        121 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 EGNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE 212

Qy        181 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA 272

Qy        241 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLF 332

Qy        301 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        333 KGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPK 392

Qy        361 LRDEF 365
              |||||
Db        393 LRDEF 397


RESULT 4
AAM78377
ID   AAM78377 standard; protein; 417 AA.
XX
AC   AAM78377;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 1039.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y;
PI   Ma Y,  Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK51510.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 20; Page 3276-3277; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 417 AA;

  Query Match             94.4%;  Score 1867.5;  DB 4;  Length 417;
  Best Local Similarity   96.2%;  Pred. No. 3.8e-193;
  Matches  350;  Conservative    3;  Mismatches    6;  Indels    5;  Gaps    1;

Qy          7 DPAGYLLYCPCM-----GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
```

```
           | | :  : | :     ||||||||||||||||||||||||||||||||||||||||||||||||||
Db.         54 DMATWPTHAPSVEEGGQGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 113

Qy          62 VSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKE 121
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         114 VSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKE 173

Qy         122 GNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEE 181
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         174 GNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEE 233

Qy         182 HRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMAS 241
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         234 HRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMAS 293

Qy         242 PQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFK 301
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         294 PQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFK 353

Qy         302 GKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKL 361
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         354 GKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKL 413

Qy         362 RDEF 365
               ||||
Db         414 RDEF 417
```

```
RESULT 5
AAM79361
ID   AAM79361 standard; protein; 417 AA.
XX
AC   AAM79361;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 3007.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y;
PI   Ma Y,  Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK52494.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 20; Page 229-230; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 417 AA;
```

```
  Query Match           93.8%;  Score 1855.5;  DB 4;  Length 417;
  Best Local Similarity  95.6%;  Pred. No. 7.6e-192;
  Matches  348;  Conservative    4;  Mismatches    7;  Indels    5;  Gaps    1;

Qy           7 DPAGYLLYCPCM-----GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
               | | :  : | :     ||||||||||||||||||||:|||| ||||||||||||||||
Db          54 DMATWPTHAPSVEEGGQGRFGNQADHFLGSLAFAKLLNRSLAVPSWIEYQHHKPPFTNLH 113

Qy          62 VSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKE 121
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         114 VSYQKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKE 173

Qy         122 GNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEE 181
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         174 GNPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEE 233

Qy         182 HRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMAS 241
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         234 HRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMAS 293

Qy         242 PQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFK 301
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         294 PQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFK 353

Qy         302 GKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKL 361
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         354 GKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKL 413

Qy         362 RDEF 365
               ||||
Db         414 RDEF 417
```

```
RESULT 6
AAW80577
ID   AAW80577 standard; protein; 343 AA.
XX
AC   AAW80577;
XX
DT   16-DEC-1998  (first entry)
XX
DE   Partial human sequence of unknown function from a myeblast cell line.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS   Homo sapiens.
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
```

```
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
DR   N-PSDB; AAV65634.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example 1; Page 62-68; 90pp; English.
XX
CC   This represents a published partial human sequence of unknown function
CC   from a myeblast cell line. The invention provides a human heart O-
CC   fucosyltransferase enzyme that can glycosylate an epidermal growth factor
CC   (EGF) domain of a polypeptide with an activated O-fucose residue.
CC   Inhibitors of O-fucosyltransferase, e.g. mutants with increased affinity
CC   for the EFG domains, are used in diagnosis and treatment of conditions
CC   associated with overexpression of O-fucosyltransferase, to promote
CC   survival of sensory (retinal) neurons. Probes based on EGF domain
CC   polypeptide are used to detect gene amplification and expression. The
CC   expression can also be determined at the protein level using antibodies
CC   specific for O-fucosyltransferase
XX
SQ   Sequence 343 AA;

  Query Match             93.2%;  Score 1844;  DB 2;  Length 343;
  Best Local Similarity   100.0%;  Pred. No. 1e-190;
  Matches  343;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         23 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRVISL 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRVISL 60

Qy         83 EDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELF 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELF 120

Qy        143 TGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQ 202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQ 180

Qy        203 IHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLP 262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 IHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLP 240

Qy        263 DLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILG 322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 DLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILG 300

Qy        323 QADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 365
              |||||||||||||||||||||||||||||||||||||||||||
Db        301 QADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 343


RESULT 7
ABB63790
ID   ABB63790 standard; protein; 402 AA.
XX
AC   ABB63790;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 18162.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL07893.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 18162; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 402 AA;

  Query Match             42.0%;  Score 831;  DB 4;  Length 402;
  Best Local Similarity   44.7%;  Pred. No. 1.1e-80;
  Matches  167;  Conservative   62;  Mismatches  119;  Indels   26;  Gaps    9;

Qy          7 DPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQK 66
              || ||| |||||||||||||||||||||| ||||| :|||:||:  :    :  | :
Db         28 DPNGYLTYCPCMGRFGNQADHFLGSLAFAKALNRTLILPPWVEYR--RGELRSRQVPFNT 85

Qy         67 YFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFE-----VAAQRSPDKKTCPMKE 121
              ||::|||: |||||:: |||  ||   ||  :||::|::        :  |||  |  |:
Db         86 YFEVEPLKEYHRVITMADFMWHLADDIWPESERVSFCYKERYSLQQEKNDPDKPNCHAKD 145

Qy        122 GNPFGPFWDQFHVSFNKSELFTGISFSASYREQ---WSQRFSPKEHPVLALPGAPAQFPV 178
              ||||||||| ||: | :|| :  : |   :  :   |  ::  : :||||  |||| |||
Db        146 GNPFGPFWDTFHIDFVRSEFYAPLHFDVHHSNEAAKWQTKYPAESYPVLAFTGAPASFPV 205

Qy        179 LEEHRPLQKYMVWSDEMVKTGEAQIHAHLVR-PYVGIHLRIGSDWKNACAMLKDGTAGSH 237
              |:  ||:|: ||    :  :  |   | |  ::||||| | ||  ||  :||      |
Db        206 QLENCKLQRYLQWSQRYREASKDFIREQLPRGAFLGIHLRNGIDWVRACEHVKD---SQH 262

Qy        238 FMASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSL-------DAQSVYVATDSE 290
              |||||:||     | |   :|:|  : | | :|  ::::       : :||:||:||
Db        263 LFASPQCLGYKNERGA-LYPELCMPSKEAIIRQLKRTIKNVRQTQPDNEIKSVFVASDSN 321

Qy        291 SYVPELQ-QLFKGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRP 349
              : ||   | :  : |  |  :   :|| ||||::|||||:||::|| |||||: | |
Db        322 HMIGELNTALSRMGISVHKLPEDDPYLDLAILGQSNHFIGNCISSYSAFEKRERDVHGFP 381

Qy        350 SSFFGMDRPPKLRD 363
              | |:|    || :|
Db        382 SYFWGF---PKEKD 392


RESULT 8
AAE03110
ID   AAE03110 standard; protein; 165 AA.
XX
AC   AAE03110;
XX
```

```
DT   10-AUG-2001  (first entry)
XX
DE   Human gene 18 encoded secreted protein HMWEY26, SEQ ID NO:144.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder; gene therapy;
KW   endocrine disorder; infection; wound healing; vulnerary; cell culture;
KW   chemotaxis; food additive; chromosome 3; binding partner identification.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .1
FT                   /label= Signal_peptide
FT   Protein         2. .165
FT                   /note= "Mature secreted protein"
FT   Misc-difference 2
FT                   /label= Unknown
FT                   /note= "Encoded by ANA"
FT   Misc-difference 4
FT                   /label= Unknown
FT                   /note= "Encoded by NCA"
FT   Misc-difference 5
FT                   /label= Unknown
FT                   /note= "Encoded by TGA"
FT   Misc-difference 148
FT                   /label= Unknown
FT                   /note= "Encoded by AGS"
XX
PN   WO200132676-A1.
XX
PD   10-MAY-2001.
XX
PF   25-OCT-2000; 2000WO-US029365.
XX
PR   29-OCT-1999;   99US-0162237P.
PR   21-JUL-2000; 2000US-0219666P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Shi Y,  Olsen HS,  Soppet DR;
XX
DR   WPI; 2001-328773/34.
DR   N-PSDB; AAD07629.
XX
PT   Nucleic acids encoding 25 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. Gaucher's disease,
PT   Alzheimer's disease, Scimitar syndrome, Creutzfeldt-Jacob disease,
PT   diabetes mellitus and multiple sclerosis.
XX
PS   Claim 11; Page 486-487; 546pp; English.
XX
CC   AAD07571-AAD07645 represent cDNAs corresponding to 25 human secreted
CC   protein genes, and AAE03052-AAE03126 represent the proteins they encode.
CC   AAE03127-AAE03150 represent human secreted protein fragments. The genes
CC   and their corresponding secreted proteins are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   25 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinsons's disease), cognitive disorders, schizophrenia, asthma, skin
CC   disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein of the invention
XX
SQ   Sequence 165 AA;

  Query Match             40.9%;  Score 809;  DB 4;  Length 165;
  Best Local Similarity   96.9%;  Pred. No. 6.6e-79;
  Matches  155;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        206 HLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLK 265
              |  | | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          6 HFARTYPGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLK 65

Qy        266 EIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILGQAD 325
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 EIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILGQAD 125

Qy        326 HFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 365
              ||||||||||||||||||| |||||||||||||||||
Db        126 HFIGNCVSSFTAFVKRERDLQGXPSSFFGMDRPPKLRDEF 165


RESULT 9
ABG63619
ID   ABG63619 standard; protein; 165 AA.
XX
AC   ABG63619;
XX
DT   27-AUG-2002  (first entry)
XX
DE   Human albumin fusion protein #294.
XX
KW   Albumin fusion protein; therapeutic protein X; human albumin; HA;
KW   human serum albumin; HSA; cancer; reproductive disorder;
KW   digestive disorder; immune disorder; endocrine disorder;
KW   haematopoietic disorder; neural disorder; connective disorder;
KW   cytostatic; antiinfertility; antiinflammatory; antiulcer;
KW   immunomodulator; anti-HIV; antidiabetic; haemostatic; nootropic;
KW   neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;
KW   osteopathic; antiarthritic.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
PN   WO200177137-A1.
XX
PD   18-OCT-2001.
XX
PF   12-APR-2001; 2001WO-US011988.
XX
PR   12-APR-2000; 2000US-0229358P.
PR   25-APR-2000; 2000US-0199384P.
PR   21-DEC-2000; 2000US-0256931P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Haseltine WA;
XX
DR   WPI; 2002-010886/01.
XX
PT   New fusion protein for treating disease e.g. diabetes comprises an
```

PT albumin fused to a therapeutic protein.
XX
PS Claim 1; Page 755-756; 2102pp; English.
XX
CC The present invention relates to albumin fusion proteins comprising a
CC therapeutic protein X and human albumin (HA, also known as human serum
CC albumin, HSA). The proteins are useful for treating a disease or disorder
CC that may be modulated by therapeutic protein X. The albumin extends the
CC shelf-life of protein X, and may increase its biological in vitro/in vivo
CC activity. The protein is useful for treating and diagnosing disorders
CC such as cancer, reproductive disorders, digestive disorders (e.g. Crohn's
CC disease, ulcerative colitis), immune disorders (e.g. acquired
CC immunodeficiency syndrome, AIDS), endocrine disorders (e.g. diabetes),
CC haematopoietic disorders, neural disorders (e.g. Alzheimer's,
CC Parkinson's, Creutzfeldt-Jacob disease, encephalomyelitis, meningitis,
CC schizophrenia), and connective disorders (e.g. osteoporosis, arthritis).
CC ABG63326-ABG65518 represent albumin fusion proteins of the invention
XX
SQ Sequence 165 AA;

Query Match 40.9%; Score 809; DB 5; Length 165;
Best Local Similarity 96.9%; Pred. No. 6.6e-79;
Matches 155; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 206 HLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLK 265
Db 6 HFARTYPGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLK 65

Qy 266 EIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILGQAD 325
Db 66 EIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILGQAD 125

Qy 326 HFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 365
Db 126 HFIGNCVSSFTAFVKRERDLQGXPSSFFGMDRPPKLRDEF 165

RESULT 10
ADL76884
ID ADL76884 standard; protein; 165 AA.
XX
AC ADL76884;
XX
DT 20-MAY-2004 (first entry)
XX
DE Albumin fusion protein related therapeutic protein X, SEQ ID No 366.
XX
KW albumin fusion protein; cytostatic; antianaemic; antiarthritic;
KW antiasthmatic; anti-HIV; immunosuppressive; antiinflammatory;
KW antipsoriatic; antibacterial; osteopathic; dermatological; antigout;
KW immunomodulator; antiarrhythmic; cardiant; nootropic; antilipaemic;
KW nephrotropic; uropathic; neuroprotective; antiparkinsonian; tranquilizer;
KW antidiabetic; anabolic; hypertensive; vulnerary; gene therapy; cancer;
KW reproductive system disorder; therapeutic protein.
XX
OS Unidentified.
XX
PN US2004010134-A1.
XX
PD 15-JAN-2004.
XX
PF 12-APR-2001; 2001US-00833245.
XX
PR 12-APR-2000; 2000US-0229358P.
PR 25-APR-2000; 2000US-0199384P.
PR 21-DEC-2000; 2000US-0256931P.
XX
PA (ROSE/) ROSEN C A.
PA (HASE/) HASELTINE W A.
XX
PI Rosen CA, Haseltine WA;
XX
DR WPI; 2004-090519/09.
XX
PT New albumin fusion proteins, useful for diagnosing, treating, preventing
PT or ameliorating diseases or disorders e.g. cancer, anemia, arthritis,
PT asthma, inflammatory bowel disease or Alzheimer's disease.
XX
PS Disclosure; SEQ ID NO 366; 279pp; English.
XX
CC The invention relates to a novel albumin fusion protein. The invention
CC further relates to: a composition comprising the albumin fusion protein
CC and a pharmaceutical carrier; a kit comprising the composition of the
CC albumin fusion protein formula; a method of treating a disease or
CC disorder in a patient comprising the step of administering the albumin
CC fusion protein; a method of treating a patient with a disease or disorder
CC that is modulated by Therapeutic protein: X, or its fragment or variant;
CC a method of extending the shelf life of Therapeutic protein: X, or its
CC fragment or variant; a nucleic acid molecule comprising a polynucleotide
CC sequence encoding the albumin fusion protein; a vector comprising the
CC nucleic acid molecule of the albumin fusion protein; and a host cell
CC comprising the nucleic acid molecule of the albumin fusion protein. The
CC albumin fusion protein and its compositions have the following
CC activities: cytostatic, antianaemic, antiarthritic, antiasthmatic, anti-
CC HIV, immunosuppressive, antiinflammatory, antipsoriatic, antibacterial,
CC osteopathic, dermatological, antigout, immunomodulator, antiarrhythmic,
CC cardiant, nootropic, antilipaemic, nephrotropic, uropathic,
CC neuroprotective, antiparkinsonian, tranquilizer, antidiabetic, anabolic,
CC hypertensive, and vulnerary. The albumin fusion protein nucleic acid may
CC be used in gene therapy to treat disorders. The albumin fusion protein is
CC useful for diagnosing, treating, preventing or ameliorating diseases or
CC disorders comprising indication: Y. The diseases or disorders include:
CC cancer (e.g. leukaemia, colon, bone, breast, liver or lung cancer),
CC immune or haematopoietic diseases (e.g. anaemia, Hodgkin's disease, acute
CC lymphocytic anaemia, multiple myeloma, arthritis, asthma, AIDS,
CC autoimmune disease, inflammatory bowel disease, psoriasis or Lyme
CC disease), reproductive system disorders (e.g. prostatitis, inguinal
CC hernia, varicocele, penile carcinoma, ovarian adenocarcinoma or Sertoli-
CC leydig tumours), musculoskeletal diseases (e.g. giant cell tumours,
CC Paget's disease, systemic lupus erythematosus, gout, muscular dystrophy
CC or cachexia), cardiovascular disease (e.g. rhabdomyomas, heart disease,
CC arrhythmia, cardiac arrest, heat valve disease, hypernatraemia or
CC hyponatraemia), mixed foetal diseases (e.g. foetal alcohol syndrome,
CC Down's syndrome, Patau syndrome, Turner's syndrome, Apert syndrome or Tay
CC -Sachs disease), excretory diseases (e.g. urinary incontinence, urinary
CC tract infections or renal disorders), neural or sensory disease (e.g.
CC Alzheimer's disease, Parkinson's disease, cerebral malaria, meningitis,
CC cerebellar ataxia, attention deficit disorder, autism or obsessive
CC compulsive disorder), respiratory disease (e.g. emphysema, lung cancer or
CC occupational lung disease), endocrine diseases (e.g. diabetes, Addison's
CC disease or glomerulonephritis), digestive diseases (e.g. portal
CC hypertension, irritable bowel disease, gastric atrophy or pancreatitis)
CC or connective tissue or epithelial diseases (e.g. Crohn's disease,
CC scleroderma, wound healing or epidermolysis bullosa). This sequence
CC represents a therapeutic protein X relating to the albumin fusion protein
CC of the invention. The sequence listing data for this specification was
CC dowloaded from the USPTO website.
XX
SQ Sequence 165 AA;

Query Match 40.9%; Score 809; DB 8; Length 165;
Best Local Similarity 96.9%; Pred. No. 6.6e-79;
Matches 155; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 206 HLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLK 265
Db 6 HFARTYPGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLK 65

Qy 266 EIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILGQAD 325
Db 66 EIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILGQAD 125

Qy 326 HFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF 365
Db 126 HFIGNCVSSFTAFVKRERDLQGXPSSFFGMDRPPKLRDEF 165

```
RESULT 11
AAW80574
ID   AAW80574 standard; peptide; 61 AA.
XX
AC   AAW80574;
XX
DT   16-DEC-1998  (first entry)
XX
DE   N-terminal amino acid sequence of human heart O-fucosyltransferase.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS   Homo sapiens.
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Claim 4; Page 69; 90pp; English.
XX
CC   This represents a the N-terminal sequence of the human heart O-
CC   fucosyltransferase that can glycosylate an epidermal growth factor (EGF)
CC   domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC   O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 61 AA;

  Query Match             17.8%;  Score 353;  DB 2;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-30;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 12
AAW80572
ID   AAW80572 standard; peptide; 61 AA.
XX
AC   AAW80572;
XX
DT   16-DEC-1998  (first entry)
XX
DE   N-terminal amino acid sequence of CHO O-fucosyltransferase.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 15
FT                   /note= "unknown"
FT   Misc-difference 17
FT                   /note= "unknown"
FT   Misc-difference 38
FT                   /note= "unknown"
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Claim 5; Page 69; 90pp; English.
XX
CC   This represents the N-terminal sequence of CHO O-fucosyltransferase. The
CC   enzyme can glycosylate an epidermal growth factor (EGF) domain of a
CC   polypeptide with an activated O-fucose residue. Inhibitors of O-
CC   fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 61 AA;

  Query Match             14.9%;  Score 295;  DB 2;  Length 61;
  Best Local Similarity   91.5%;  Pred. No. 8.6e-24;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||
Db          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 13
ABG95669
ID   ABG95669 standard; protein; 328 AA.
XX
AC   ABG95669;
XX
DT   06-DEC-2002  (first entry)
XX
DE   Human nucleic acid associated protein, NAAP26, Incyte 7503717CD1.
XX
KW   Human; nucleic acid-associated protein; NAAP; gene therapy; microarray;
KW   proliferative disorder; developmental disorder; cardiovascular disorder;
KW   neurological disorder; autoimmune disorder; inflammatory disorder;
KW   atherosclerosis; bursitis; cirrhosis; hepatitis; psoriasis; cancer;
KW   lymphoma; melanoma; brain cancer; breast cancer; Cushing's syndrome;
KW   Alzheimer's disease; AIDS; Creutzfeldt-Jakob disease;
KW   acquired immunodeficiency syndrome; Goodpasture's syndrome;
```

```
KW   Crohn's disease; multiple sclerosis.
XX
OS   Homo sapiens.
XX
PN   WO200274913-A2.
XX
PD   26-SEP-2002.
XX
PF   14-MAR-2002; 2002WO-US007869.
XX
PR   16-MAR-2001; 2001US-0276857P.
PR   19-APR-2001; 2001US-0285489P.
PR   19-APR-2001; 2001US-0285556P.
PR   04-MAY-2001; 2001US-0288646P.
PR   04-MAY-2001; 2001US-0288700P.
PR   10-MAY-2001; 2001US-0290510P.
PR   11-MAY-2001; 2001US-0290369P.
PR   16-NOV-2001; 2001US-0332426P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Hillman JL,  Baughn MR,  Swarnakar A,  Yue H,  Elliott VS,  Burford N;
PI   Ding L,  Tang YT,  Lee SY,  Azimzai Y,  Walia NK,  Gietzen KJ;
PI   Griffin JA,  Lal PG,  Yang J,  Borowsky ML,  Richardson TW,  Yue H;
PI   Becha S,  Forsythe IJ,  Jones KA,  Warren BA,  Thangavelu K;
PI   Honchell CD,  Jolley HE,  Hafalia AJA,  Ring HZ;
XX
DR   WPI; 2002-723448/78.
DR   N-PSDB; ABS73886.
XX
PT   New human nucleic acid-associated proteins and polynucleotides, useful
PT   for diagnosing, treating or preventing cardiovascular, neurological,
PT   autoimmune or inflammatory disorders, e.g. atherosclerosis, Alzheimer's
PT   or AIDS.
XX
PS   Claim 1; Page 193-194; 219pp; English.
XX
CC   The invention relates to an isolated human nucleic acid-associated
CC   protein (NAAP), comprising the protein sequences appearing as ABG95644-
CC   ABG95669, or an immunogenic or biologically active fragment. Also
CC   included are the NAAP encoding polynucleotides, a recombinant
CC   polynucleotide comprising a promoter sequence operably linked to the NAAP
CC   polynucleotide, a cell transformed with the recombinant NAAP
CC   polynucleotide, an NAAP transgenic organism, an anti-NAAP antibody,
CC   screening compounds as modulators or ant/agonists of NAAP or modulators
CC   of NAAP polynucleotide expression, a microarray comprising 30-60
CC   nucleotides of the NAAP polynucleotides and generating an expression
CC   profile of a sample that contains NAAP polynucleotides. The polypeptide,
CC   polynucleotide and agonist are useful for treating a disease or condition
CC   associated with decreased expression of functional NAAP. The antagonist
CC   is useful for treating a disease or condition associated with
CC   overexpression of functional NAAP. The antibody that specifically binds
CC   to the polypeptide is useful for diagnosing a condition or disease
CC   associated with the expression of NAAP. These diseases include
CC   proliferative, development, cardiovascular, neurological, or autoimmune
CC   or inflammatory disorders. These polypeptides, polynucleotides, agonists
CC   and antagonists may also be used for preventing these diseases. These
CC   disease or conditions associated with functional NAAP expression also
CC   includes atherosclerosis, bursitis, cirrhosis, hepatitis, psoriasis,
CC   cancers (e.g. lymphoma, melanoma, brain cancer or breast cancer),
CC   Cushing's syndrome, Alzheimer's disease, acquired immunodeficiency
CC   syndrome (AIDS), Creutzfeldt-Jakob disease, Goodpasture's syndrome,
CC   Crohn's disease, or multiple sclerosis and many other diseases and
CC   conditions listed in the specification. The present sequence represents
CC   an NAAP protein
XX
SQ   Sequence 328 AA;

  Query Match             8.3%;  Score 164.5;  DB 5;  Length 328;
  Best Local Similarity   21.7%;  Pred. No. 1.6e-08;
  Matches   85;  Conservative   41;  Mismatches  115;  Indels  151;  Gaps   15;

Qy          3 AGSWDPAG-----------------------YLLY--CPCMGRFGNQADHFLGSLAFAK 36
              | || ||                       ||||   |  | |  : | ::   :  |
Db         13 AVSWPPASASGQEFWPGQSAADILSGAASRRRYLLYDVNPPEG-FNLRRDVYIRIASLLK 71

Qy         37 LLNRT----LAVPPWIEYQH-HKPPFTNLHVSYQKYFKLEPLQAYHRVISLEDFMEKLAP 91
              | :|    | :|||    |   |    : : : ::| |  |     ||  | |: :
Db         72 TLLKTEEWVLVLPPWGRLYHWQSPDIHQVRIPWSEFFDLPSLNKNIPVIEYEQFIAESG- 130

Qy         92 THWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFNKSELFTGISFSASY 151
                                              ||| || :|              ||
Db        131 ----------------------------------GPFIDQVYV-------------LQSY 143

Qy        152 REQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMV-----WSDEMVKTGEAQIHAH 206
               | | :     |  |             : ||    ::     |    || | |
Db        144 AEGWKE--GTWEEKV-------------DERPCIDQLLYFQEDWMKMKVKLGSA----- 182

Qy        207 LVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAAPLTMTMCLPDLKE 266
              |  ||:|:||            ||   |                      :| |:
Db        183 LGGPYLGVHLR-----------RKDFIWGHR--------------------QDVPSLEG 210

Qy        267 IQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKPE----------VAQ 315
              | ::  :::     |:||||:     | ::| |   ::|  :|             ||
Db        211 AVRKIRSLMKTHRLDKVFVATDAVR--KEYEELKKLLPEMVRFEPTWEELELYKDGGVAI 268

Qy        316 VDLYILGQADHFIGNCVSSFTAFVKRERDLQG 347
              :| :|   |  |||  ||:|:  :  ||:: |
Db        269 IDQWICAHARFFIGTSVSTFSFRIHEEREILG 300


RESULT 14
ABB69800
ID   ABB69800 standard; protein; 490 AA.
XX
AC   ABB69800;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 36192.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL13903.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 36192; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
```

CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 490 AA;

Query Match 8.1%; Score 159.5; DB 4; Length 490;
Best Local Similarity 22.4%; Pred. No. 1.1e-07;
Matches 92; Conservative 55; Mismatches 159; Indels 105; Gaps 19;

```
Qy          9 AGYLLYCPCMGR-FGNQADHFLGSLAFAKLLNR-------TLAVPPWIEYQH-HKPPFTN 59
              | |:||   :   |  : | ::    | : | |        | :|||    | |
Db         71 AVYILYDVNISEGFNLRRDVYIRMAVFVRRLQRRRRFRHVRLVLPPWPRLYHWHSQGLQQ 130

Qy         60 LHVSYQKYFKLEPLQAYHRVISLEDFM--EKLAPTHWPPEKRVAYCFEV----------- 106
               : :  :| |  |: |  |:  |:|:  ::|     |   | : | :
Db        131 SGLPWSHFFDLASLRRYAPVLDYEEFLAEQRLFGNPGAPLVHVGHAFRLQHYEVMLEQGI 190

Qy        107 ---AAQRSPDKKTCPMKEGN-PFGPFWDQFHVSFNKSELFTGISFSAS-------YREQW 155
                :|  ||   |  ||:   ||   |  :   :   | : |  |         ||
Db        191 FRDKFERVTDK---PCSEGSLSGGPLLQQAELRVGR---FHCVRFQGSAGLLEKLLREAI 244

Qy        156 SQRFSPKEH----PVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQIHAHLVRPY 211
               :  :  |         ||  |     || :|        | ||          |    |
Db        245 DEDTAGPEDVDDMRTYALLSAET---VLHDH--------WGDE---------HFWQARRS 284

Qy        212 VGIHLR---IGSDWKNACAMLKDGTA-------------------GSHFMASPQCVGYSR 249
               :    |   : :|::       | :|                  | :  |  :   : |
Db        285 MRFARRLEQVAADFRRQALDTTDASAGVQRPAMWELERPKRNAKGGDYLCAHLRRGDFVR 344

Qy        250 STAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESY-VPELQQLFKGKVKVVS 308
              |  |        | ||   : ||  :|  :  :|::|||:  | : ||::||  : ::|
Db        345 SRDA------TTPTLKAAAQQVKQLLRGFNMTTVFLATDATPYELMELKELFY-RFRLVH 397

Qy        309 LKPE------------VAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQG 347
                ||            || ||  :   | :|:|   |:||  :  ||:: |
Db        398 FAPESNVQRRELKDGGVAVVDQLVCAYARYFVGTYESTFTYRIYEEREILG 448
```

RESULT 15
ADS96504
ID ADS96504 standard; protein; 490 AA.
XX
AC ADS96504;
XX
DT 02-DEC-2004 (first entry)
XX
DE Drosophila melanogaster protein, SEQ ID 125.
XX
KW Insecticide; Antiparasitic; Antihelminthic.
XX
OS Drosophila melanogaster.
XX
PN WO2004039999-A2.
XX
PD 13-MAY-2004.
XX
PF 08-AUG-2003; 2003WO-US024982.
XX
PR 30-OCT-2002; 2002US-0422377P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Stam L, Kamdar KP, Spana E, Bachmann J;
XX
DR WPI; 2004-376203/35.
DR N-PSDB; ADS96503.
XX
PT Identifying a compound that inhibits the activity of a protein for
PT Drosophila viability for use e.g., as insecticidal agent by expressing in
PT a recombinant host a DNA molecule to produce a protein essential for
PT Drosophila viability.
XX
PS Claim 1; SEQ ID NO 125; 57pp; English.

XX
CC The present invention relates to a method for identifying a compound that
CC inhibits the activity of a protein essential for Drosophila viability.
CC The method comprises: (a) expressing in a recombinant host a DNA sequence
CC encoding a protein essential for Drosophila viability; (b) testing
CC compounds suspected of having the ability to inhibit the activity of the
CC protein expressed in (a); and identifying a compound tested in (b) that
CC inhibits the activity of the protein. The method is useful in identifying
CC a compound that inhibits the activity of a protein essential for
CC Drosophila viability for use as insecticidal, ectoparasiticidal,
CC antiparasitic, antihelminthic or acaracidal agent. The present sequence
CC is one such protein essential for Drosophila viability. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 490 AA;

Query Match 8.1%; Score 159.5; DB 8; Length 490;
Best Local Similarity 22.4%; Pred. No. 1.1e-07;
Matches 92; Conservative 55; Mismatches 159; Indels 105; Gaps 19;

```
Qy          9 AGYLLYCPCMGR-FGNQADHFLGSLAFAKLLNR-------TLAVPPWIEYQH-HKPPFTN 59
              | |:||   :   |  : | ::    | : | |        | :|||    | |
Db         71 AVYILYDVNISEGFNLRRDVYIRMAVFVRRLQRRRRFRHVRLVLPPWPRLYHWHSQGLQQ 130

Qy         60 LHVSYQKYFKLEPLQAYHRVISLEDFM--EKLAPTHWPPEKRVAYCFEV----------- 106
               : :  :| |  |: |  |:  |:|:  ::|     |   | : | :
Db        131 SGLPWSHFFDLASLRRYAPVLDYEEFLAEQRLFGNPGAPLVHVGHAFRLQHYEVMLEQGI 190

Qy        107 ---AAQRSPDKKTCPMKEGN-PFGPFWDQFHVSFNKSELFTGISFSAS-------YREQW 155
                :|  ||   |  ||:   ||   |  :   :   | : |  |         ||
Db        191 FRDKFERVTDK---PCSEGSLSGGPLLQQAELRVGR---FHCVRFQGSAGLLEKLLREAI 244

Qy        156 SQRFSPKEH----PVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTGEAQIHAHLVRPY 211
               :  :  |         ||  |     || :|        | ||          |    |
Db        245 DEDTAGPEDVDDMRTYALLSAET---VLHDH--------WGDE---------HFWQARRS 284

Qy        212 VGIHLR---IGSDWKNACAMLKDGTA-------------------GSHFMASPQCVGYSR 249
               :    |   : :|::       | :|                  | :  |  :   : |
Db        285 MRFARRLEQVAADFRRQALDTTDASAGVQRPAMWELERPKRNAKGGDYLCAHLRRGDFVR 344

Qy        250 STAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESY-VPELQQLFKGKVKVVS 308
              |  |        | ||   : ||  :|  :  :|::|||:  | : ||::||  : ::|
Db        345 SRDA------TTPTLKAAAQQVKQLLRGFNMTTVFLATDATPYELMELKELFY-RFRLVH 397

Qy        309 LKPE------------VAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQG 347
                ||            || ||  :   | :|:|   |:||  :  ||:: |
Db        398 FAPESNVQRRELKDGGVAVVDQLVCAYARYFVGTYESTFTYRIYEEREILG 448
```

Search completed: October 25, 2005, 15:30:37
Job time : 236.341 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 08:18:39 ; Search time 320.031 Seconds
(without alignments)
1866.199 Million cell updates/sec

Title: US-09-774-954-2
Perfect score: 1979
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....QGRPSSFFGMDRPPKLRDEF 365

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105433_6415/app_query.fasta_1.917
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954_@CGN_1_1_116_@runat_25102005_105433_6415 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1979	100.0	1100	3	US-08-978-741-16	Sequence 16, Appl
2	1979	100.0	1100	3	US-09-333-729A-16	Sequence 16, Appl
3	1979	100.0	1300	3	US-08-978-741-4	Sequence 4, Appli
4	1979	100.0	1300	3	US-09-333-729A-6	Sequence 6, Appli
5	1979	100.0	1514	3	US-08-978-741-1	Sequence 1, Appli
6	1979	100.0	1514	3	US-09-333-729A-2	Sequence 2, Appli
7	1979	100.0	11284	3	US-08-978-741-5	Sequence 5, Appli
8	1844	93.2	5009	3	US-08-978-741-7	Sequence 7, Appli
9	1844	93.2	5009	3	US-09-333-729A-8	Sequence 8, Appli
10	837	42.3	1320	4	US-09-270-767-14353	Sequence 14353, A
11	343	17.3	722	4	US-09-270-767-30507	Sequence 30507, A
c 12	186	9.4	998	4	US-09-270-767-28905	Sequence 28905, A
c 13	186	9.4	2034	4	US-09-270-767-13022	Sequence 13022, A
c 14	120	6.1	19227	4	US-09-949-016-12127	Sequence 12127, A
c 15	120	6.1	19228	4	US-09-949-016-16285	Sequence 16285, A
16	100.5	5.1	10259	4	US-09-949-016-13785	Sequence 13785, A
17	97.5	4.9	24741	4	US-09-949-016-15547	Sequence 15547, A
c 18	96.5	4.9	3899	4	US-09-902-540-573	Sequence 573, App
c 19	96.5	4.9	28374	4	US-09-949-016-17508	Sequence 17508, A
c 20	96.5	4.9	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c 21	94.5	4.8	4086	4	US-09-614-221A-478	Sequence 478, App
c 22	94.5	4.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 23	94	4.7	1505	4	US-09-291-299A-2	Sequence 2, Appli
c 24	94	4.7	2768	4	US-09-774-528-426	Sequence 426, App
25	94	4.7	74881	4	US-09-949-016-15545	Sequence 15545, A
26	94	4.7	74914	4	US-09-949-016-12286	Sequence 12286, A
c 27	93.5	4.7	36542	4	US-09-949-016-12149	Sequence 12149, A
c 28	93.5	4.7	36544	4	US-09-949-016-13434	Sequence 13434, A
29	92.5	4.7	3816	4	US-09-540-236-1820	Sequence 1820, Ap
30	92.5	4.7	6558	4	US-09-491-356C-7	Sequence 7, Appli
31	92.5	4.7	63563	4	US-09-596-002-33	Sequence 33, Appl
32	92	4.6	77626	4	US-09-949-016-12608	Sequence 12608, A
33	91.5	4.6	2229	4	US-09-214-904-1	Sequence 1, Appli
34	91.5	4.6	145812	4	US-09-949-016-15698	Sequence 15698, A
35	91	4.6	893	4	US-09-270-767-11464	Sequence 11464, A
36	91	4.6	5591	4	US-09-949-016-12044	Sequence 12044, A
37	91	4.6	5592	4	US-09-949-016-16387	Sequence 16387, A
c 38	91	4.6	10357	3	US-08-961-527-191	Sequence 191, App
c 39	91	4.6	12665	4	US-09-949-016-16388	Sequence 16388, A
c 40	90.5	4.6	1499	4	US-09-509-908-1	Sequence 1, Appli
41	90	4.5	20870	4	US-09-949-016-16017	Sequence 16017, A
42	90	4.5	35100	2	US-08-770-379-17	Sequence 17, Appl
43	90	4.5	35100	3	US-08-757-669A-17	Sequence 17, Appl
44	90	4.5	35100	3	US-09-230-371A-17	Sequence 17, Appl
c 45	89.5	4.5	2501	4	US-09-614-891-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-978-741-16
; Sequence 16, Application US/08978741 16
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:

```
;      LENGTH:  1100 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-16

Alignment Scores:
Pred. No.:                2.04e-234      Length:          1100
Score:                    1979.00        Matches:         365
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-774-954-2 (1-365) x US-08-978-741-16 (1-1100)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1081 CTGCGGGACGAGTTC 1095

RESULT 2
US-09-333-729A-16
; Sequence 16, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 16
;   LENGTH: 1100
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-16

Alignment Scores:
Pred. No.:                2.04e-234      Length:          1100
Score:                    1979.00        Matches:         365
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-774-954-2 (1-365) x US-09-333-729A-16 (1-1100)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360
```

```
Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1081 CTGCGGGACGAGTTC 1095

RESULT 3
US-08-978-741-4
; Sequence 4, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1300 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-4

Alignment Scores:
Pred. No.:                2.69e-234      Length:          1300
Score:                    1979.00        Matches:         365
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              100.00%        Indels:          0
DB:                       3              Gaps:            0

US-09-774-954-2 (1-365) x US-08-978-741-4 (1-1300)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1216 CTGCGGGACGAGTTC 1230

RESULT 4
US-09-333-729A-6
; Sequence 6, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 6
;   LENGTH: 1300
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid insert.
US-09-333-729A-6

Alignment Scores:
Pred. No.:              2.69e-234      Length:          1300
Score:                  1979.00        Matches:         365
Percent Similarity:     100.00%        Conservative:    0
Best Local Similarity:  100.00%        Mismatches:      0
Query Match:            100.00%        Indels:          0
DB:                     3              Gaps:            0

US-09-774-954-2 (1-365) x US-09-333-729A-6 (1-1300)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 375

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 435

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 495

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        496 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 555

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        556 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 615

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        616 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 675

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        676 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 735

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        736 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 795

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        796 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 855

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        856 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 915

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        916 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 975

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        976 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 1035

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1036 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 1095

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1096 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1155

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1156 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1215

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1216 CTGCGGGACGAGTTC 1230

RESULT 5
US-08-978-741-1
; Sequence 1, Application US/08978741
; Patent No. 6100076
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1514 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-1

Alignment Scores:
Pred. No.:                3.47e-234       Length:          1514
Score:                    1979.00         Matches:         365
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       3               Gaps:            0

US-09-774-954-2 (1-365) x US-08-978-741-1 (1-1514)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1081 CTGCGGGACGAGTTC 1095

RESULT 6
US-09-333-729A-2
; Sequence 2, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 2
;   LENGTH: 1514
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-2

Alignment Scores:
Pred. No.:                3.47e-234       Length:          1514
```

```
Score:                  1979.00       Matches:       365
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     3             Gaps:          0

US-09-774-954-2 (1-365) x US-09-333-729A-2 (1-1514)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 240

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 300

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 360

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 420

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 480

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAA 540

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 600

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 660

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 720

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 780

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 840

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 900

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 960

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 1020

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 1080

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       1081 CTGCGGGACGAGTTC 1095

RESULT 7
US-08-978-741-5
; Sequence 5, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11284 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-5

Alignment Scores:
Pred. No.:              9.61e-233     Length:        11284
Score:                  1979.00       Matches:       365
Percent Similarity:     100.00%       Conservative:  0
Best Local Similarity:  100.00%       Mismatches:    0
Query Match:            100.00%       Indels:        0
DB:                     3             Gaps:          0

US-09-774-954-2 (1-365) x US-08-978-741-5 (1-11284)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy         61 HisValSerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIle 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4416 CATGTGTCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATC 4475

Qy         81 SerLeuGluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgVal 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4476 AGCTTGGAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTG 4535

Qy        101 AlaTyrCysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLys 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4536 GCATACTGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAG 4595

Qy        121 GluGlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGlu 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4596 GAAGGAAACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAG 4655

Qy        141 LeuPheThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSer 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4656 CTTTTTACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCT 4715

Qy        161 ProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGlu 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4716 CCAAAGGAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAG 4775

Qy        181 GluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGlu 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4776 GAACACAGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAG 4835

Qy        201 AlaGlnIleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySer 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4836 GCCCAGATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCT 4895

Qy        221 AspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAla 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4896 GACTGGAAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCC 4955

Qy        241 SerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCys 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4956 TCTCCGCAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGC 5015

Qy        261 LeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAla 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5016 CTGCCTGACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCC 5075

Qy        281 GlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPhe 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5076 CAGTCGGTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTC 5135

Qy        301 LysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIle 320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5136 AAAGGGAAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATC 5195

Qy        321 LeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLys 340
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5196 CTCGGCCAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAG 5255

Qy        341 ArgGluArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLys 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5256 CGGGAGCGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAG 5315

Qy        361 LeuArgAspGluPhe 365
              |||||||||||||||
Db       5316 CTGCGGGACGAGTTC 5330

RESULT 8
US-08-978-741-7
; Sequence 7, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5009 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-7

Alignment Scores:
Pred. No.:                1.24e-216       Length:          5009
Score:                    1844.00         Matches:         343
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              93.18%          Indels:          0
DB:                       3               Gaps:            0

US-09-774-954-2 (1-365) x US-08-978-741-7 (1-5009)

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 61

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisVal 62
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGTG 121

Qy         63 SerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeu 82
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTTG 181

Qy         83 GluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyr 102
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATAC 241

Qy        103 CysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGly 122
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGGA 301

Qy        123 AsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPhe 142
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 AACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTTT 361

Qy        143 ThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerProLys 162
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 ACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAAG 421
```

```
Qy        163 GluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGluHis 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 GAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACAC 481

Qy        183 ArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAlaGln 202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCAG 541

Qy        203 IleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTrp 222
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 ATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTGG 601

Qy        223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 AAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCG 661

Qy        243 GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro 262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 CAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCT 721

Qy        263 AspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSer 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 GACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCG 781

Qy        283 ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLysGly 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 GTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGG 841

Qy        303 LysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGly 322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 AAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGC 901

Qy        323 GlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArgGlu 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 CAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAG 961

Qy        343 ArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLysLeuArg 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 CGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGG 1021

Qy        363 AspGluPhe 365
              |||||||||
Db       1022 GACGAGTTC 1030

RESULT 9
US-09-333-729A-8
; Sequence 8, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 8
;   LENGTH: 5009
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-8

Alignment Scores:
Pred. No.:                1.24e-216      Length:          5009
Score:                    1844.00        Matches:         343
Percent Similarity:       100.00%        Conservative:    0
Best Local Similarity:    100.00%        Mismatches:      0
Query Match:              93.18%         Indels:          0
DB:                       3              Gaps:            0

US-09-774-954-2 (1-365) x US-09-333-729A-8 (1-5009)
```

```
Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 61

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisVal 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCATGTG 121

Qy         63 SerTyrGlnLysTyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeu 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 TCCTACCAGAAGTACTTCAAGCTGGAGCCCCTCCAGGCTTACCATCGGGTCATCAGCTTG 181

Qy         83 GluAspPheMetGluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyr 102
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 GAGGATTTCATGGAGAAGCTGGCACCCACCCACTGGCCCCCTGAGAAGCGGGTGGCATAC 241

Qy        103 CysPheGluValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGluGly 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 TGCTTTGAGGTGGCAGCCCAGCGAAGCCCAGATAAGAAGACGTGCCCCATGAAGGAAGGA 301

Qy        123 AsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeuPhe 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 AACCCCTTTGGCCCATTCTGGGATCAGTTTCATGTGAGTTTCAACAAGTCGGAGCTTTTT 361

Qy        143 ThrGlyIleSerPheSerAlaSerTyrArgGluGlnTrpSerGlnArgPheSerProLys 162
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 ACAGGCATTTCCTTCAGTGCTTCCTACAGAGAACAATGGAGCCAGAGATTTTCTCCAAAG 421

Qy        163 GluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProValLeuGluGluHis 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 GAACATCCGGTGCTTGCCCTGCCAGGAGCCCCAGCCCAGTTCCCCGTCCTAGAGGAACAC 481

Qy        183 ArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThrGlyGluAlaGln 202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AGGCCACTACAGAAGTACATGGTATGGTCAGACGAAATGGTGAAGACGGGAGAGGCCCAG 541

Qy        203 IleHisAlaHisLeuValArgProTyrValGlyIleHisLeuArgIleGlySerAspTrp 222
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 ATTCATGCCCACCTTGTCCGGCCCTATGTGGGCATTCATCTGCGCATTGGCTCTGACTGG 601

Qy        223 LysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHisPheMetAlaSerPro 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 AAGAACGCCTGTGCCATGCTGAAGGACGGGACTGCAGGCTCGCACTTCATGGCCTCTCCG 661

Qy        243 GlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMetThrMetCysLeuPro 262
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 CAGTGTGTGGGCTACAGCCGCAGCACAGCGGCCCCCCTCACGATGACTATGTGCCTGCCT 721

Qy        263 AspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSerLeuAspAlaGlnSer 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 GACCTGAAGGAGATCCAGAGGGCTGTGAAGCTCTGGGTGAGGTCGCTGGATGCCCAGTCG 781

Qy        283 ValTyrValAlaThrAspSerGluSerTyrValProGluLeuGlnGlnLeuPheLysGly 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 GTCTACGTTGCTACTGATTCCGAGAGTTATGTGCCTGAGCTCCAACAGCTCTTCAAAGGG 841

Qy        303 LysValLysValValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGly 322
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 AAGGTGAAGGTGGTGAGCCTGAAGCCTGAGGTGGCCCAGGTCGACCTGTACATCCTCGGC 901

Qy        323 GlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAlaPheValLysArgGlu 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 CAAGCCGACCACTTTATTGGCAACTGTGTCTCCTCCTTCACTGCCTTTGTGAAGCGGGAG 961

Qy        343 ArgAspLeuGlnGlyArgProSerSerPhePheGlyMetAspArgProProLysLeuArg 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 CGGGACCTCCAGGGGAGGCCGTCTTCTTTCTTCGGCATGGACAGGCCCCCTAAGCTGCGG 1021

Qy        363 AspGluPhe 365
              |||||||||
Db       1022 GACGAGTTC 1030
```

```
RESULT 10
US-09-270-767-14353
; Sequence 14353, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14353
;   LENGTH: 1320
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-14353

Alignment Scores:
Pred. No.:              4.55e-93      Length:          1320
Score:                  837.00        Matches:         168
Percent Similarity:     61.50%        Conservative:    62
Best Local Similarity:  44.92%        Mismatches:      118
Query Match:            42.29%        Indels:          26
DB:                     4             Gaps:            9

US-09-774-954-2 (1-365) x US-09-270-767-14353 (1-1320)

Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGlnAlaAsp 26
              |||||||  |||||||||  ||||||||||||||||||||||||||||||||||||||||
Db         75 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCGAC 134

Qy         27 HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProPro 46
              ||||||||||||||||||||||||||||||  ||||||||||||||  :::|||||||
Db        135 CACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCG 194

Qy         47 TrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHisValSerTyrGlnLys 66
              |||:::||||||:::   :::          :::       |||   :::
Db        195 TGGGTGGAGTATCGT------AGGGGTGAACTGCGATCCCGTCAGGTACCGTTCAACACA 248

Qy         67 TyrPheLysLeuGluProLeuGlnAlaTyrHisArgValIleSerLeuGluAspPheMet 86
              ||||||:::::|||||||||:::  ||||||||||||||:::::   |||||||||
Db        249 TATTTTGAAGTGGAGCCCCTGAAGGAATACCATCGCGTCATCACCATGGCAGATTTCATG 308

Qy         87 GluLysLeuAlaProThrHisTrpProProGluLysArgValAlaTyrCysPheGlu--- 105
                 ||||||          ||||||      :::|||||:::::|||:::::
Db        309 TGGCACCTGGCCGACGACATTTGGCCAGAATCGGAGCGAGTGTCATTTTGCTACAAGGAA 368

Qy        106 ------------ValAlaAlaGlnArgSerProAspLysLysThrCysProMetLysGlu 121
                        :::      |||||||||       |||      |||:::
Db        369 CGATATAGCCTTCAGCAGGAGAAGAACGATCCAGACAAGCCCAATTGCCACGCCAAGGAT 428

Qy        122 GlyAsnProPheGlyProPheTrpAspGlnPheHisValSerPheAsnLysSerGluLeu 141
              ||||||||||||||||||||||||||  ||||||:::   |||   :::|||||||
Db        429 GGCAATCCTTTTGGTCCCTTTTGGGACACTTTTCACATAGACTTTGTGCGGTCAGAGTTC 488

Qy        142 PheThrGlyIleSerPheSerAlaSerTyrArgGluGln---------TrpSerGlnArg 158
              :::     :::   |||          :::     :::         |||      :::
Db        489 TATGCGCCACTTCATTTTGATGTGCATCATAGCAACGAGGCTGCCAAGTGGCAGACCAAA 548

Qy        159 PheSerProLysGluHisProValLeuAlaLeuProGlyAlaProAlaGlnPheProVal 178
              :::      :::   :::||||||||||||      ||||||||||   |||||||||
Db        549 TATCCTGCAGAATCATATCCCGTACTCGCGTTCACCGGAGCTCCGGCTAGTTTTCCTGTT 608

Qy        179 LeuGluGluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMetValLysThr 198
                 |||:::     ||||||:::|||:::  ||||||          :::
Db        609 CAGCTAGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGTATAGGGAAGCA 668

Qy        199 GlyGluAlaGlnIleHisAlaHisLeuValArg---ProTyrValGlyIleHisLeuArg 217
                 :::      |||         |||   |||      ::::::|||||||||||||||
Db        669 TCTAAGGATTTCATCCGAGAGCAGTTGCCTCGGGGTGCCTTTTTGGGCATTCATCTGCGC 728

Qy        218 IleGlySerAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAlaGlySerHis 237
                    |||   ||||||      ||||||      :::||||||          |||
Db        729 AACGGTATCGATTGGGTGAGAGCCTGTGAGCACGTCAAGGAT---------AGCCAGCAT 779

Qy        238 PheMetAlaSerProGlnCysValGlyTyrSerArgSerThrAlaAlaProLeuThrMet 257
                 ||||||||||||||:::||||||          |||   |||
Db        780 CTGTTTGCCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTGCA---CTCTACCCG 836

Qy        258 ThrMetCysLeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrpValArgSer 277
                 :::|||:::|||      :::   |||   |||   :::|||      :::::::::
Db        837 GAGCTCTGCATGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGAACCATTAAGAAC 896

Qy        278 Leu--------------------AspAlaGlnSerValTyrValAlaThrAspSerGlu 290
                 :::                 :::   :::||||||:::||||||:::|||||
Db        897 GTGCGCCAAACTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCGTCAGACTCCAAT 956

Qy        291 SerTyrValProGluLeuGln---GlnLeuPheLysGlyLysValLysValValSerLeu 309
                 :::   ||||||        |||   :::       :::   |||      |||
Db        957 CACATGATTGGTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGTGTGCACAAGCTG 1016

Qy        310 LysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAspHisPheIleGly 329
                 :::         :::|||||   ||||||||||||:::::||||||||||||
Db       1017 ACGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACCACTTTATCGGC 1076

Qy        330 AsnCysValSerSerPheThrAlaPheValLysArgGluArgAspLeuGlnGlyArgPro 349
              ||||||:::||||||:::::||||||||||||||||||||||:::   |||   |||
Db       1077 AACTGTATATCCTCTTACTCGGCATTCGTAAAAAGGGAACGAGATGTGCACGGTTTTCCA 1136

Qy        350 SerSerProPheGlyMetAspArgProProLysLeuArgAsp 363
              |||   |||:::|||            ||||||   :::|||
Db       1137 TCGTACTTCTGGGGATTC---------CCCAAGGAAAAGGAT 1169

RESULT 11
US-09-270-767-30507
; Sequence 30507, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30507
;   LENGTH: 722
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-30507

Alignment Scores:
Pred. No.:              2.05e-32      Length:          722
Score:                  343.00        Matches:         80
Percent Similarity:     57.87%        Conservative:    34
Best Local Similarity:  40.61%        Mismatches:      67
Query Match:            17.33%        Indels:          16
DB:                     4             Gaps:            6

US-09-774-954-2 (1-365) x US-09-270-767-30507 (1-722)

Qy        176 PheProValLeuGluGluHisArgProLeuGlnLysTyrMetValTrpSerAspGluMet 195
              |||||||||      |||:::      |||||||:::|||:::   ||||||
Db          2 TTTCCTGTTCAGCTAGAGAACTGCAAGCTGCAGCGCTACTTGCAGTGGAGTCAACGGTAT 61

Qy        196 ValLysThrGlyGluAlaGlnIleHisAlaHisLeuValArg---ProTyrValGlyIle 214
                 :::      :::       |||   |||      ::::::||||||
Db         62 AGGGAAGCATCTAAGGATTTCATCCGAGAGCAGTTGCCTCGGGGTGCCTTTTTGGGCATT 121

Qy        215 HisLeuArgIleGlySerAspTrpLysAsnAlaCysAlaMetLeuLysAspGlyThrAla 234
              |||||||||   |||   ||||||      ||||||      :::||||||
Db        122 CATCTGCGCAACGGTATCGATTGGGTGAGAGCCTGTGAGCACGTCAAGGAT--------- 172

Qy        235 GlySerHisPheMetAlaSerProGlnCysValGlyTyrSerArgSerThrAlaAlaPro 254
                    |||       ||||||||||||||:::||||||                |||
```

```
Db        173 AGCCAGCATCTGTTTGCCTCGCCGCAGTGCTTGGGCTATAAAAATGAACGTGGTGCA--- 229

Qy        255 LeuThrMetThrMetCysLeuProAspLeuLysGluIleGlnArgAlaValLysLeuTrp 274
              |||          :::|||:::|||       :::   |||   |||   :::|||
Db        230 CTCTACCCGGAGCTCTGCATGCCCTCCAAGGAGGCGATCATCCGCCAGCTAAAGAGAACC 289

Qy        275 ValArgSerLeu---------------------AspAlaGlnSerValTyrValAlaThr 287
              ::::::::::::                   :::   :::||||||:::||||||:::
Db        290 ATTAAGAACGTGCGCCAAACTCAGCCGGACAACGAAATCAAATCAGTTTTCGTGGCGTCA 349

Qy        288 AspSerGluSerTyrValProGluLeuGln---GlnLeuPheLysGlyLysValLysVal 306
              ||||||         :::   ||||||          |||   :::       :::   |||
Db        350 GACTCCAATCACATGATTGGTGAACTAAACACGGCCCTTAGTCGCATGGGCATCAGTGTG 409

Qy        307 ValSerLeuLysProGluValAlaGlnValAspLeuTyrIleLeuGlyGlnAlaAspHis 326
                  |||      :::       :::||||||   ||||||||||||:::::::|||
Db        410 CACAAGCTGACGGAGGATGATCCTTACCTGGACTTGGCCATTCTCGGACAGTCGAACCAC 469

Qy        327 PheIleGlyAsnCysValSerSerPheThrAlaPheValLysArgGluArgAspLeuGln 346
              ||||||||||||||:::||||||::::::||||||||||||||||||||||||:::
Db        470 TTTATCGGCAACTGTATATCCTCTTACTCGGCATTCGTAAAAAGGGAACGAGATGTGCAC 529

Qy        347 GlyArgProSerSerPhePheGlyMetAspArgProProLysLeuArgAsp 363
              |||   ||||||   |||:::|||          ||||||   :::|||
Db        530 GGTTTTCCATCGTACTTCTGGGGATTC---------CCCAAGGAAAAGGAT 571

RESULT 12
US-09-270-767-28905/c
; Sequence 28905, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 28905
;   LENGTH: 998
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-28905

Alignment Scores:
Pred. No.:                9.08e-13        Length:         998
Score:                    186.00          Matches:        41
Percent Similarity:       67.44%          Conservative:   17
Best Local Similarity:    47.67%          Mismatches:     24
Query Match:              9.40%           Indels:         4
DB:                       4               Gaps:           2

US-09-774-954-2 (1-365) x US-09-270-767-28905 (1-998)

Qy        279 AspAlaGlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGln--- 297
              :::   :::||||||:::||||||:::||||||         :::   ||||||
Db        975 GAAATCAAATCAGTTTTCGTGGCGTCAGACTCCAATCACATGATTGGTGAACTAAACACG 916

Qy        298 GlnLeuPheLysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAsp 317
                  |||      :::       :::   |||       |||       :::       :::|||
Db        915 GCCCTTAGTCGCATGGGCATCAGTGTGCACAAGCTGACGGAGGATGATCCTTACCTGGAC 856

Qy        318 LeuTyrIleLeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAla 337
              |||   ||||||||||||:::::::||||||||||||||||||:::||||||:::::|||
Db        855 TTGGCCATTCTCGGACAGTCGAACCACTTTATCGGCAACTGTATATCCTCTTACTCGGCA 796

Qy        338 PheValLysArgGluArgAspLeuGlnGlyArgProSerSerPheGlyMetAspArg 357
              ||||||||||||||||||:::   |||   ||||||   |||:::|||
Db        795 TTCGTAAAAAGGGAACGAGATGTGCACGGTTTTCCATCGTACTTCTGGGGATTC------ 742

Qy        358 ProProLysLeuArgAsp 363
                 ||||||   :::|||
Db        741 ---CCCAAGGAAAAGGAT 727

RESULT 13
US-09-270-767-13022/c
; Sequence 13022, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13022
;   LENGTH: 2034
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-13022

Alignment Scores:
Pred. No.:                2.95e-12        Length:         2034
Score:                    186.00          Matches:        41
Percent Similarity:       67.44%          Conservative:   17
Best Local Similarity:    47.67%          Mismatches:     24
Query Match:              9.40%           Indels:         4
DB:                       4               Gaps:           2

US-09-774-954-2 (1-365) x US-09-270-767-13022 (1-2034)

Qy        279 AspAlaGlnSerValTyrValAlaThrAspSerGluSerTyrValProGluLeuGln--- 297
              :::   :::||||||:::||||||:::||||||         :::   ||||||
Db       2011 GAAATCAAATCAGTTTTCGTGGCGTCAGACTCCAATCACATGATTGGTGAACTAAACACG 1952

Qy        298 GlnLeuPheLysGlyLysValLysValValSerLeuLysProGluValAlaGlnValAsp 317
                  |||   :::       :::   |||       |||       :::       :::|||
Db       1951 GCCCTTAGTCGCATGGGCATCAGTGTGCACAAGCTGACGGAGGATGATCCTTACCTGGAC 1892

Qy        318 LeuTyrIleLeuGlyGlnAlaAspHisPheIleGlyAsnCysValSerSerPheThrAla 337
              |||   ||||||||||||:::::::||||||||||||||||||:::||||||:::::|||
Db       1891 TTGGCCATTCTCGGACAGTCGAACCACTTTATCGGCAACTGTATATCCTCTTACTCGGCA 1832

Qy        338 PheValLysArgGluArgAspLeuGlnGlyArgProSerSerPheGlyMetAspArg 357
              ||||||||||||||||||:::   |||   ||||||   |||:::|||
Db       1831 TTCGTAAAAAGGGAACGAGATGTGCACGGTTTTCCATCGTACTTCTGGGGATTC------ 1778

Qy        358 ProProLysLeuArgAsp 363
                 ||||||   :::|||
Db       1777 ---CCCAAGGAAAAGGAT 1763

RESULT 14
US-09-949-016-12127/c
; Sequence 12127, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12127
;   LENGTH: 19227
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12127
```

Alignment Scores:

Pred. No.:	0.0175	Length:	19227
Score:	120.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	95.00%	Mismatches:	0
Query Match:	6.06%	Indels:	0
DB:	4	Gaps:	0

US-09-774-954-2 (1-365) x US-09-949-016-12127 (1-19227)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       1718 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGTAAG 1659
```

RESULT 15
US-09-949-016-16285/c
; Sequence 16285, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16285
; LENGTH: 19228
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16285

Alignment Scores:

Pred. No.:	0.0175	Length:	19228
Score:	120.00	Matches:	19
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	95.00%	Mismatches:	0
Query Match:	6.06%	Indels:	0
DB:	4	Gaps:	0

US-09-774-954-2 (1-365) x US-09-949-016-16285 (1-19228)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       1718 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGTAAG 1659
```

Search completed: October 26, 2005, 15:24:38
Job time : 354.031 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:21:23 ; Search time 35.1971 Seconds
(without alignments)
887.482 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGYLLYXPXMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	320	98.2	61	1	OFU1_CRIGR	P83337 cricetulus
2	305	93.6	393	1	OFU1_MOUSE	Q91zw2 mus musculu
3	305	93.6	393	2	Q8C8R4	Q8c8r4 mus musculu
4	302	92.6	395	2	Q6EV70	Q6ev70 rattus norv
5	295	90.5	388	1	OFU1_HUMAN	Q9h488 homo sapien
6	295	90.5	388	1	OFU1_PANTR	Q6ev69 pan troglod
7	288	88.3	351	2	Q7YRE6	Q7yre6 bos taurus
8	288	88.3	391	2	Q7YRE7	Q7yre7 bos taurus
9	286	87.7	378	2	Q7YRZ3	Q7yrz3 sus scrofa
10	273	83.7	391	2	Q6EV66	Q6ev66 bos taurus
11	272	83.4	380	2	Q8AWB4	Q8awb4 gallus gall
12	250	76.7	380	2	Q640S0	Q640s0 xenopus tro
13	250	76.7	396	2	Q6EV71	Q6ev71 xenopus tro
14	249	76.4	395	2	Q7T028	Q7t028 brachydanio
15	247	75.8	384	2	Q70AG7	Q70ag7 fugu rubrip
16	247	75.8	387	2	Q6EV65	Q6ev65 tetraodon n
17	234	71.8	380	2	Q8AXS8	Q8axs8 xenopus lae
18	177	54.3	402	1	OFU1_DROME	Q9v6x7 drosophila
19	177	54.3	402	2	Q86SA7	Q86sa7 drosophila
20	177	54.3	402	2	Q659S1	Q659s1 drosophila
21	171	52.5	399	2	Q7QHS7	Q7qhs7 anopheles g
22	170	52.1	381	2	Q6EV67	Q6ev67 ciona savig
23	164	50.3	402	2	Q6EV68	Q6ev68 drosophila
24	156	47.9	385	2	Q659R9	Q659r9 onchocerca
25	150.5	46.2	353	2	Q659S0	Q659s0 bombyx mori
26	150	46.0	364	2	Q8MSR1	Q8msr1 drosophila
27	142	43.6	381	1	OFU1_CAEEL	Q18014 caenorhabdi
28	136	41.7	434	2	Q659S2	Q659s2 caenorhabdi
29	77.5	23.8	602	2	Q9L0H6	Q9l0h6 streptomyce
30	72.5	22.2	601	2	Q82DG5	Q82dg5 streptomyce
31	68.5	21.0	569	2	Q9CCV1	Q9ccv1 mycobacteri
32	68.5	21.0	585	2	Q59530	Q59530 mycobacteri
33	68	20.9	375	2	Q76B90	Q76b90 prunus aviu
34	67.5	20.7	354	1	OPSI_ASTFA	P51474 astyanax fa
35	67.5	20.7	376	1	FUT1_RAT	Q10980 rattus norv
36	67.5	20.7	538	2	Q7XC59	Q7xc59 oryza sativ
37	67.5	20.7	538	2	Q9AUT5	Q9aut5 oryza sativ
38	66.5	20.4	376	1	FUT1_MOUSE	O09160 mus musculu
39	66.5	20.4	377	2	P97327	P97327 mus musculu
40	66.5	20.4	377	2	Q920W4	Q920w4 mus spicile
41	66.5	20.4	377	2	Q920W5	Q920w5 mus musculu
42	66.5	20.4	377	2	Q9JL28	Q9jl28 mus musculu
43	66.5	20.4	767	1	FTSK_STRPN	P64166 streptococc
44	66.5	20.4	767	1	FTSK_STRR6	P64167 streptococc
45	64	19.6	832	2	Q7YTA5	Q7yta5 leishmania

ALIGNMENTS

RESULT 1
OFU1_CRIGR
ID OFU1_CRIGR STANDARD; PRT; 61 AA.
AC P83337;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE GDP-fucose protein O-fucosyltransferase 1 (EC 2.4.1.221) (Peptide O-
DE fucosyltransferase) (O-FucT-1) (Fragment).
GN Name=POFUT1;
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE.
RX MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;
RA Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,
RA Haltiwanger R.S.;
RT "Modification of epidermal growth factor-like repeats with O-fucose:
RT molecular cloning and expression of a novel GDP-fucose protein O-
RT fucosyltransferase.";
RL J. Biol. Chem. 276:40338-40345(2001).
RN [2]
RP FUNCTION.
RX MEDLINE=93364082; PubMed=8358148;
RA Harris R.J., Spellman M.W.;
RT "O-linked fucose and other post-translational modifications unique to
RT EGF modules.";
RL Glycobiology 3:219-224(1993).
RN [3]
RP FUNCTION.
RX MEDLINE=97175972; PubMed=9023546;
RA Wang Y., Lee G.F., Kelley R.F., Spellman M.W.;
RT "Identification of a GDP-L-fucose:polypeptide fucosyltransferase and
RT enzymatic addition of O-linked fucose to EGF domains.";
RL Glycobiology 6:837-842(1996).
RN [4]
RP FUNCTION, AND GLYCOSYLATION.
RC TISSUE=Ovary;
RX MEDLINE=98192601; PubMed=9525914; DOI=10.1074/jbc.273.14.8112;
RA Wang Y., Spellman M.W.;
RT "Purification and characterization of a GDP-fucose:polypeptide
RT fucosyltransferase from Chinese hamster ovary cells.";
RL J. Biol. Chem. 273:8112-8118(1998).
CC -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC O-glycosidic linkage to a conserved serine or threonine residue in
CC EGF domains.
CC -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC -!- COFACTOR: Manganese. Other divalent cations increase activity:
CC calcium, cobalt, cadmium, magnesium and nickel.
CC -!- PATHWAY: Glycosylation.

```
CC   -!- PTM: N-glycosylated.
CC   -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
KW   Direct protein sequencing; Fucose metabolism; Glycoprotein;
KW   Glycosyltransferase; Manganese; Transferase.
FT   NON_TER      61     61
SQ   SEQUENCE   61 AA;  6951 MW;  E9507AE60018F23A CRC64;

  Query Match             98.2%;  Score 320;  DB 1;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 3.2e-35;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RLAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RLAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 2
OFU1_MOUSE
ID   OFU1_MOUSE     STANDARD;      PRT;   393 AA.
AC   Q91ZW2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)
DE   (Peptide O-fucosyltransferase) (O-FucT-1).
GN   Name=Pofut1;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ; TISSUE=Liver;
RX   MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;
RA   Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,
RA   Haltiwanger R.S.;
RT   "Modification of epidermal growth factor-like repeats with O-fucose:
RT   molecular cloning and expression of a novel GDP-fucose protein O-
RT   fucosyltransferase.";
RL   J. Biol. Chem. 276:40338-40345(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N-3; TISSUE=Mammary gland;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC       O-glycosidic linkage to a conserved serine or threonine residue in
CC       EGF domains (By similarity).
CC   -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC       beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC   -!- COFACTOR: Manganese (By similarity).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF375885; AAL09577.1; -.
DR   EMBL; BC046295; AAH46295.1; -.
DR   MGD; MGI:2153207; Pofut1.
DR   GO; GO:0030173; C:integral to Golgi membrane; ISS.
DR   GO; GO:0008417; F:fucosyltransferase activity; ISS.
DR   GO; GO:0001525; P:angiogenesis; IMP.
DR   GO; GO:0007507; P:heart development; IMP.
DR   GO; GO:0007399; P:neurogenesis; IMP.
DR   GO; GO:0007219; P:Notch signaling pathway; IMP.
DR   GO; GO:0006493; P:O-linked glycosylation; ISS.
DR   GO; GO:0001756; P:somitogenesis; IMP.
KW   Fucose metabolism; Glycoprotein; Glycosyltransferase; Manganese;
KW   Signal; Transferase.
FT   SIGNAL        1     30       Potential.
FT   CHAIN        31    393       GDP-fucose protein O-fucosyltransferase
FT                                1.
FT   CARBOHYD     67     67       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    165    165       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   393 AA;  44688 MW;  D982104E95E5CF3B CRC64;

  Query Match             93.6%;  Score 305;  DB 1;  Length 393;
  Best Local Similarity   91.8%;  Pred. No. 2.3e-32;
  Matches   56;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 RLAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNL 60
              | |||||||||||| | ||||||||||||||||||||  ||||||||||||||||||||
Db         29 RSAGSWDLAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 88

Qy         61 H 61
              |
Db         89 H 89


RESULT 3
Q8C8R4
ID   Q8C8R4      PRELIMINARY;      PRT;   393 AA.
AC   Q8C8R4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Mus musculus adult retina cDNA, RIKEN full-length enriched library,
DE   clone:A930028F21 product:protein O-fucosyltransferase 1, full insert
DE   sequence.
GN   Name=Pofut1;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
```

```
RL   Nature 409:685-690(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK044629; BAC32009.1; -.
DR   MGD; MGI:2153207; Pofut1.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0001525; P:angiogenesis; IMP.
DR   GO; GO:0006004; P:fucose metabolism; TAS.
DR   GO; GO:0007507; P:heart development; IMP.
DR   GO; GO:0007399; P:neurogenesis; IMP.
DR   GO; GO:0007219; P:Notch signaling pathway; IMP.
DR   GO; GO:0001756; P:somitogenesis; IMP.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   393 AA;  44704 MW;  1CA94CB30B85D6DC CRC64;

  Query Match             93.6%;  Score 305;  DB 2;  Length 393;
  Best Local Similarity   91.8%;  Pred. No. 2.3e-32;
  Matches   56;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          1 RLAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNL 60
              | ||||||||||||| | |||||||||||||||||||  ||||||||||||||||||||
Db         29 RSAGSWDLAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 88

Qy         61 H 61
              |
Db         89 H 89


RESULT 4
Q6EV70
ID   Q6EV70      PRELIMINARY;      PRT;   395 AA.
AC   Q6EV70;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Protein-O-fucosyltransferase 1.
GN   Name=fut12;
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar;
RX   PubMed=12966037; DOI=10.1093/glycob/cwg113;
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RA   Oriol R.;
RT   "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT   fucosyltransferases and alpha6-fucosyltransferases: phylogeny and
RT   identification of conserved peptide motifs.";
RL   Glycobiology 13:1C-5C(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar;
RA   Martinez-Duncker I., Oriol R., Mollicone R.;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ781499; CAH03711.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   395 AA;  44983 MW;  2A85521E22F38920 CRC64;

  Query Match             92.6%;  Score 302;  DB 2;  Length 395;
  Best Local Similarity   90.2%;  Pred. No. 5.9e-32;
  Matches   55;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 RLAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNL 60
              | ||||||||||||| | |||||||||||||||||||  ||||||||||||||||||||
Db         31 RAGGSWDLAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 90

Qy         61 H 61
              |
Db         91 H 91


RESULT 5
OFU1_HUMAN
ID   OFU1_HUMAN     STANDARD;      PRT;   388 AA.
AC   Q9H488; Q14685; Q9BW76;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)
DE   (Peptide O-fucosyltransferase) (O-FucT-1).
GN   Name=POFUT1; Synonyms=KIAA0180;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND INDUCTION.
RC   TISSUE=Heart;
RX   MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;
RA   Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,
RA   Haltiwanger R.S.;
RT   "Modification of epidermal growth factor-like repeats with O-fucose:
RT   molecular cloning and expression of a novel GDP-fucose protein O-
RT   fucosyltransferase.";
RL   J. Biol. Chem. 276:40338-40345(2001).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain;
RX   MEDLINE=22158633; PubMed=12168954;
```

```
RA   Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT   "Construction of expression-ready cDNA clones for KIAA genes: manual
RT   curation of 330 KIAA cDNA clones.";
RL   DNA Res. 9:99-106(2002).
RN   [3]
RP   SEQUENCE OF 46-388 FROM N.A. (ISOFORM 1).
RC   TISSUE=Bone marrow;
RX   MEDLINE=96281124; PubMed=8724849;
RA   Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. V.
RT   The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT   analysis of cDNA clones from human cell line KG-1.";
RL   DNA Res. 3:17-24(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21638749; PubMed=11780052; DOI=10.1038/414865a;
RA   Deloukas P., Matthews L.H., Ashurst J.L., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA   Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA   Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA   Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA   Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA   Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA   Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA   Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA   Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [6]
RP   CHARACTERIZATION.
RX   MEDLINE=97175972; PubMed=9023546;
RA   Wang Y., Lee G.F., Kelley R.F., Spellman M.W.;
RT   "Identification of a GDP-L-fucose:polypeptide fucosyltransferase and
RT   enzymatic addition of O-linked fucose to EGF domains.";
RL   Glycobiology 6:837-842(1996).
CC   -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC       O-glycosidic linkage to a conserved serine or threonine residue in
CC       EGF domains.
CC   -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC       beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC   -!- COFACTOR: Manganese.
CC   -!- PATHWAY: Glycosylation.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9H488-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9H488-2; Sequence=VSP_001809;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Highly expressed in heart, brain, placenta,
CC       lung, liver, skeletal muscle, kidney and pancreas.
CC   -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF375884; AAL09576.1; -.
DR   EMBL; D80002; BAA11497.2; -.
DR   EMBL; AL121897; CAC16424.1; -.
DR   EMBL; BC000582; AAH00582.1; -.
DR   Genew; HGNC:14988; POFUT1.
DR   H-InvDB; HIX0015724; -.
DR   MIM; 607491; -.
DR   GO; GO:0005737; C:cytoplasm; ISS.
DR   GO; GO:0030173; C:integral to Golgi membrane; NAS.
DR   GO; GO:0046922; F:peptide-O-fucosyltransferase activity; TAS.
DR   GO; GO:0009790; P:embryonic development; NAS.
DR   GO; GO:0007219; P:Notch signaling pathway; ISS.
DR   GO; GO:0016266; P:O-glycan processing; TAS.
DR   GO; GO:0045449; P:regulation of transcription; NAS.
DR   GO; GO:0006445; P:regulation of translation; NAS.
KW   Alternative splicing; Fucose metabolism; Glycoprotein;
KW   Glycosyltransferase; Manganese; Signal; Transferase.
FT   SIGNAL        1     26       Potential.
FT   CHAIN        27    388       GDP-fucose protein O-fucosyltransferase
FT                                1.
FT   CARBOHYD     62     62       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    160    160       N-linked (GlcNAc. . .) (Potential).
FT   VARSPLIC    182    388       FSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTG
FT                                EAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA
FT                                SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQ
FT                                SVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILG
FT                                QADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRD
FT                                EF -> RENHSCVTLLFPR (in isoform 2).
FT                                /FTId=VSP_001809.
SQ   SEQUENCE   388 AA;  43955 MW;  3FACCCA434D02415 CRC64;

  Query Match             90.5%;  Score 295;  DB 1;  Length 388;
  Best Local Similarity   91.5%;  Pred. No. 5.1e-31;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||  ||||||||||||||||||||||
Db         26 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 84


RESULT 6
OFU1_PANTR
ID   OFU1_PANTR     STANDARD;      PRT;   388 AA.
AC   Q6EV69;
DT   25-OCT-2004 (Rel. 45, Created)
DT   25-OCT-2004 (Rel. 45, Last sequence update)
```

```
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)
DE   (Peptide O-fucosyltransferase) (O-FucT-1).
GN   Name=POFUT1; Synonyms=FUT12;
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=12966037; DOI=10.1093/glycob/cwg113;
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RA   Oriol R.;
RT   "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT   fucosyltransferases and alpha6-fucosyltransferases: phylogeny and
RT   identification of conserved peptide motifs.";
RL   Glycobiology 13:1C-5C(2003).
CC   -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC       O-glycosidic linkage to a conserved serine or threonine residue in
CC       EGF domains (By similarity).
CC   -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC       beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC   -!- COFACTOR: Manganese (By similarity).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AJ781500; CAH03712.1; -.
KW   Fucose metabolism; Glycoprotein; Glycosyltransferase; Manganese;
KW   Signal; Transferase.
FT   SIGNAL        1     26       Potential.
FT   CHAIN        27    388       GDP-fucose protein O-fucosyltransferase
FT                                1.
FT   CARBOHYD     62     62       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    160    160       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   388 AA;  43955 MW;  3FACCCA434D02415 CRC64;

  Query Match             90.5%;  Score 295;  DB 1;  Length 388;
  Best Local Similarity   91.5%;  Pred. No. 5.1e-31;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | |||||||||||||||||||  |||||||||||||||||||||
Db         26 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 84


RESULT 7
Q7YRE6
ID   Q7YRE6      PRELIMINARY;      PRT;   351 AA.
AC   Q7YRE6;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Protein O-fucosyltransferase 1b.
GN   Name=pofut1;
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Loriol C., Germot A., Dupuy F., Maftah A.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY344581; AAQ02333.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   351 AA;  39576 MW;  BF830F61A7296F42 CRC64;

  Query Match             88.3%;  Score 288;  DB 2;  Length 351;
  Best Local Similarity   89.7%;  Pred. No. 4e-30;
  Matches   52;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 GSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              |||| |||||| | |||||||||||||||||||  |||||||||||||||||||||:|
Db         30 GSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNVH 87


RESULT 8
Q7YRE7
ID   Q7YRE7      PRELIMINARY;      PRT;   391 AA.
AC   Q7YRE7;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Protein O-fucosyltransferase 1a.
GN   Name=pofut1;
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Loriol C., Germot A., Dupuy F., Maftah A.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY344580; AAQ02332.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   391 AA;  44307 MW;  96762AB81A2027AD CRC64;

  Query Match             88.3%;  Score 288;  DB 2;  Length 391;
  Best Local Similarity   89.7%;  Pred. No. 4.5e-30;
  Matches   52;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 GSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              |||| |||||| | |||||||||||||||||||  |||||||||||||||||||||:|
Db         30 GSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNVH 87


RESULT 9
Q7YRZ3
ID   Q7YRZ3      PRELIMINARY;      PRT;   378 AA.
AC   Q7YRZ3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Protein-O-fucosyltransferase (Fragment).
GN   Name=POFUT1;
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Martinez-Dunker I., Mollicone R., Candelier J.J., Breton C., Oriol R.;
RT   "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT   fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and
RT   identification of conserved peptide motifs.";
RL   Glycobiology 13:1c-5c(2003).
DR   EMBL; AJ567917; CAD99201.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     378    378
SQ   SEQUENCE   378 AA;  42842 MW;  1870DEB0C50713F1 CRC64;

  Query Match             87.7%;  Score 286;  DB 2;  Length 378;
  Best Local Similarity   89.8%;  Pred. No. 8e-30;
```

```
  Matches   53;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              | ||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||
Db         22 AVSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 80


RESULT 10
Q6EV66
ID   Q6EV66      PRELIMINARY;      PRT;   391 AA.
AC   Q6EV66;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Protein-O-fucosyltransferase 1.
GN   Name=fut12;
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=12966037; DOI=10.1093/glycob/cwg113;
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RA   Oriol R.;
RT   "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT   fucosyltransferases and alpha6-fucosyltransferases: phylogeny and
RT   identification of conserved peptide motifs.";
RL   Glycobiology 13:1C-5C(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Martinez-Duncker I., Oriol R., Mollicone R.;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ781503; CAH03714.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   391 AA;  44277 MW;  4E047D6BB5CDF7FA CRC64;

  Query Match             83.7%;  Score 273;  DB 2;  Length 391;
  Best Local Similarity   84.5%;  Pred. No. 4.6e-28;
  Matches   49;  Conservative    3;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          4 GSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              |||: |||||: | ||||||||||||||||||  |||||||||||||||||||:|
Db         30 GSWNPAGYLLFCPCMGRFGNQADHFLGSLAFAPLMNRTLAVPPWIEYQHHKPPFTNVH 87


RESULT 11
Q8AWB4
ID   Q8AWB4      PRELIMINARY;      PRT;   380 AA.
AC   Q8AWB4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein POFUT1.
GN   Name=POFUT1;
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Oriol R.;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ535754; CAD59738.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   380 AA;  43316 MW;  83B53E342511811B CRC64;

  Query Match             83.4%;  Score 272;  DB 2;  Length 380;
  Best Local Similarity   79.7%;  Pred. No. 6.1e-28;
  Matches   47;  Conservative    6;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||:||||:||||:   ||||||||||:||:||:|||
Db         18 AGSWDTAGYLLYCPCMGRFGNQAEHFLGALAFARALNRTLAVPPWIEYRHHRPPYTNLH 76


RESULT 12
Q640S0
ID   Q640S0      PRELIMINARY;      PRT;   380 AA.
AC   Q640S0;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein.
OS   Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8364;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Klein S., Gerhard D.S.;
RL   Submitted (SEP-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC082519; AAH82519.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   380 AA;  43570 MW;  8BB938CE776D5238 CRC64;

  Query Match             76.7%;  Score 250;  DB 2;  Length 380;
  Best Local Similarity   75.0%;  Pred. No. 5.6e-25;
  Matches   42;  Conservative    6;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          6 WDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              || :||:|| | |||||||||||||||||:  ||| ||||| | ||:||:||:|
Db         21 WDSSGYILYCPCMGRFGNQADHFLGSLAFAKMVNRTLVVPPWIVYNHHRPPYTNVH 76


RESULT 13
Q6EV71
ID   Q6EV71      PRELIMINARY;      PRT;   396 AA.
AC   Q6EV71;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Protein-O-fucosyltransferase.
GN   Name=fut12;
OS   Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
```

```
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8364;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Martinez-Duncker I., Oriol R., Mollicone R.;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ781498; CAH03710.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR   InterPro; IPR000886; ER_target_S.
DR   PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   396 AA;  45056 MW;  F90504E19507073F CRC64;

  Query Match           76.7%;  Score 250;  DB 2;  Length 396;
  Best Local Similarity   75.0%;  Pred. No. 5.8e-25;
  Matches   42;  Conservative    6;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          6 WDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              || :||:|| | |||||||||||||||||||:  ||| ||||| | ||:||:||:|
Db         37 WDSSGYILYCPCMGRFGNQADHFLGSLAFAKMVNRTLVVPPWIVYNHHRPPYTNVH 92


RESULT 14
Q7T028
ID   Q7T028      PRELIMINARY;      PRT;   395 AA.
AC   Q7T028;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   GDP-fucose protein O-fucosyltransferase 1 (EC 2.4.1.221).
GN   Name=pofut1; Synonyms=fut12;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RA   Oriol R.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ579536; CAE18459.1; -.
DR   ZFIN; ZDB-GENE-040303-2; pofut1.
DR   GO; GO:0046922; F:peptide-O-fucosyltransferase activity; IEA.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   395 AA;  45165 MW;  15BBE06172542E8A CRC64;

  Query Match           76.4%;  Score 249;  DB 2;  Length 395;
  Best Local Similarity   75.4%;  Pred. No. 7.9e-25;
  Matches   43;  Conservative    5;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          5 SWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              :||  ||:|| | ||||||| ||||||||||:  ||||||||| |:|| |||||:|
Db         33 TWDENGYILYCPCMGRFGNQVDHFLGSLAFAKMLNRTLAVPPWIVYRHHSPPFTNVH 89


RESULT 15
Q70AG7
ID   Q70AG7      PRELIMINARY;      PRT;   384 AA.
AC   Q70AG7;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Protein-O-fucosyltransferase.
GN   Name=fut12;
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetradontoidea; Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Oriol R.;
RL   Submitted (NOV-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ606070; CAE54305.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR   InterPro; IPR000886; ER_target_S.
DR   PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   384 AA;  43650 MW;  7CAA6ADC90F40032 CRC64;

  Query Match           75.8%;  Score 247;  DB 2;  Length 384;
  Best Local Similarity   76.8%;  Pred. No. 1.4e-24;
  Matches   43;  Conservative    5;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          6 WDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||  ||:|| | |||||||||||||||||||:  ||||||||| |:|| ||:||:|
Db         23 WDQNGYVLYCPCMGRFGNQADHFLGSLAFAKMLNRTLAVPPWIVYRHHVPPYTNVH 78


Search completed: October 25, 2005, 15:35:24
Job time : 36.1971 secs
```

Against SVA

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 05:59:38 ; Search time 1065.43 Seconds
(without alignments)
2179.320 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGYLLYXPXMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105432_6406/app_query.fasta_1.917
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954_@CGN_1_1_5334_@runat_25102005_105432_6406 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	305	93.6	318	6	BY778150	BY778150 BY778150
2	305	93.6	333	5	BY175344	BY175344 BY175344
3	305	93.6	352	5	BY328132	BY328132 BY328132
4	305	93.6	456	2	BB633858	BB633858 BB633858
5	305	93.6	480	2	BB851550	BB851550 BB851550
6	305	93.6	536	7	CF155036	CF155036 B0620C09-
7	305	93.6	604	2	BB618795	BB618795 BB618795
8	305	93.6	605	6	BY728018	BY728018 BY728018
9	305	93.6	611	2	BB623905	BB623905 BB623905
10	305	93.6	614	6	BY752500	BY752500 BY752500
11	305	93.6	634	2	BB631131	BB631131 BB631131
12	305	93.6	641	2	BB611933	BB611933 BB611933
13	305	93.6	650	6	BY752497	BY752497 BY752497
14	305	93.6	661	2	BB642654	BB642654 BB642654
15	305	93.6	664	6	BY728337	BY728337 BY728337
16	305	93.6	695	2	BB617910	BB617910 BB617910
17	305	93.6	915	5	BQ936634	BQ936634 AGENCOURT
18	305	93.6	947	4	BI410783	BI410783 602963556
19	305	93.6	988	5	BQ964179	BQ964179 AGENCOURT
20	305	93.6	2413	3	AK037857	AK037857 Mus muscu
21	305	93.6	2616	3	AK044629	AK044629 Mus muscu
22	302	92.6	763	7	CK366990	CK366990 AGENCOURT
23	302	92.6	3611	3	AK081059	AK081059 Mus muscu
24	298	91.4	304	5	BY322240	BY322240 BY322240
25	295	90.5	520	7	CR543310	CR543310 DKFZp459B
26	295	90.5	550	7	CN348927	CN348927 170005999
27	295	90.5	551	2	BE261136	BE261136 601152308
28	295	90.5	576	2	BE262956	BE262956 601147417
29	295	90.5	595	7	CN348926	CN348926 170005326
30	295	90.5	610	1	AL555719	AL555719 AL555719
31	295	90.5	656	7	CV029252	CV029252 7968 Full
32	295	90.5	699	1	AU137536	AU137536 AU137536
33	295	90.5	705	7	CN348925	CN348925 170005328
34	295	90.5	753	4	BG424731	BG424731 602453575
35	295	90.5	790	4	BG519639	BG519639 602578756
36	295	90.5	808	4	BG747760	BG747760 602705255
37	295	90.5	841	5	BX345989	BX345989 BX345989
38	295	90.5	889	4	BI524031	BI524031 603052202
39	295	90.5	898	2	BF315757	BF315757 601900828
40	295	90.5	919	5	BU155142	BU155142 AGENCOURT
41	295	90.5	942	7	CO774833	CO774833 ILLUMIGEN
42	295	90.5	950	4	BI199190	BI199190 602758739
43	295	90.5	955	2	BE260030	BE260030 601150556
44	295	90.5	1555	3	CR617950	CR617950 full-leng
45	295	90.5	1711	2	BF206548	BF206548 601869779

ALIGNMENTS

RESULT 1
BY778150
LOCUS BY778150 318 bp mRNA linear EST 23-MAR-2004
DEFINITION BY778150 RIKEN full-length enriched, 17.5 days embryo whole body
Mus musculus cDNA clone L930123D17 5', mRNA sequence.
ACCESSION BY778150
VERSION BY778150.1 GI:39704789
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 318)
AUTHORS Carninci,P., Waki,K., Shiraki,T., Konno,H., Shibata,K., Itoh,M.,
Aizawa,K., Arakawa,T., Ishii,Y., Sasaki,D., Bono,H., Kondo,S.,
Sugahara,Y., Saito,R., Osato,N., Fukuda,S., Sato,K., Watahiki,A.,
Hirozane-Kishikawa,T., Nakamura,M., Shibata,Y., Yasunishi,A.,
Kikuchi,N., Yoshiki,A., Kusakabe,M., Gustincich,S., Beisel,K.,
Pavan,W., Aidinis,V., Nakagawara,A., Held,W.A., Iwata,H., Kono,T.,
Nakauchi,H., Lyons,P., Wells,C., Hume,D.A., Fagiolini,M.,
Hensch,T.K., Brinkmeier,M., Camper,S., Hirota,J., Mombaerts,P.,
Muramatsu,M., Okazaki,Y., Kawai,J. and Hayashizaki,Y.
TITLE Targeting a complex transcriptome: the construction of the mouse
full-length cDNA encyclopedia
JOURNAL Genome Res. 13 (6B), 1273-1289 (2003)
MEDLINE 22703353
PUBMED 12819125
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.jp/) for further details.

FEATURES Location/Qualifiers
source 1. .318
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="L930123D17"
/tissue_type="whole body"
/dev_stage="17.5 days embryo"
/clone_lib="RIKEN full-length enriched, 17.5 days embryo whole body"

ORIGIN

Alignment Scores:

Pred. No.:	9.55e-32	Length:	318
Score:	305.00	Matches:	56
Percent Similarity:	91.80%	Conservative:	0
Best Local Similarity:	91.80%	Mismatches:	5
Query Match:	93.56%	Indels:	0
DB:	6	Gaps:	0

US-09-774-954-3 (1-61) x BY778150 (1-318)

```
Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   ||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db        115 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 174

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        175 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 234

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 294

Qy         61 His 61
              |||
Db        295 CAT 297
```

RESULT 2
BY175344
LOCUS BY175344 333 bp mRNA linear EST 10-DEC-2002
DEFINITION BY175344 RIKEN full-length enriched, NOD-derived CD11c +ve dendritic cells Mus musculus cDNA clone F630011F24 5', mRNA sequence.
ACCESSION BY175344
VERSION BY175344.1 GI:26311990
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 333)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. John Todd (Dept. of Medical Genetics Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome Trust/MRC building Addenbrookes Hospital Cambridge) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .333
/organism="Mus musculus"
/mol_type="mRNA"
/strain="NOD"
/db_xref="taxon:10090"
/clone="F630011F24"
/cell_type="NOD-derived CD11c +ve dendritic cells"
/clone_lib="RIKEN full-length enriched, NOD-derived CD11c +ve dendritic cells"

ORIGIN

Alignment Scores:

Pred. No.:	1.01e-31	Length:	333
Score:	305.00	Matches:	57
Percent Similarity:	93.44%	Conservative:	0

```
Best Local Similarity:  93.44%        Mismatches:    4
Query Match:            93.56%        Indels:        0
DB:                     5             Gaps:          0

US-09-774-954-3 (1-61) x BY175344 (1-333)

Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   ||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db         91 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 150

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        151 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGNTGAACCGC 210

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 270

Qy         61 His 61
              |||
Db        271 CAT 273

RESULT 3
BY328132
LOCUS       BY328132                352 bp    mRNA    linear   EST 11-DEC-2002
DEFINITION  BY328132 RIKEN full-length enriched, synovial fibroblasts Mus
            musculus cDNA clone L130001N19 5', mRNA sequence.
ACCESSION   BY328132
VERSION     BY328132.1  GI:26518861
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 352)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S.,
            Hirozane,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H.,
            Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R.,
            Ohno,M., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K.,
            Shiraki,T., Tagami,M., Waki,K., Watahiki,A., Muramatsu,M. and
            Hayashizaki,Y. Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Tissues were provided by Vassilis Aidinis ( Biomedical Sciences
            Research Center 'Al. Fleming' Institute of Immunology 14-16 Al.
            Fleming street 16672 Vari,Greece ) whose assistance we gratefully
            acknowledge.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES              Location/Qualifiers
     source           1. .352
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /db_xref="taxon:10090"
                      /clone="L130001N19"
                      /cell_type="synovial fibroblasts"
                      /clone_lib="RIKEN full-length enriched, synovial
                      fibroblasts"
ORIGIN

Alignment Scores:
Pred. No.:              1.08e-31      Length:        352
Score:                  305.00        Matches:       56
Percent Similarity:     91.80%        Conservative:  0
Best Local Similarity:  91.80%        Mismatches:    5
Query Match:            93.56%        Indels:        0
DB:                     5             Gaps:          0

US-09-774-954-3 (1-61) x BY328132 (1-352)

Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   ||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db        112 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 171

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        172 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 231

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        232 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 291

Qy         61 His 61
              |||
Db        292 CAT 294

RESULT 4
BB633858
LOCUS       BB633858                456 bp    mRNA    linear   EST 31-AUG-2001
DEFINITION  BB633858 RIKEN full-length enriched, adult male spinal cord Mus
            musculus cDNA clone A330084M14 5', mRNA sequence.
ACCESSION   BB633858
```

VERSION BB633858.1 GI:15400446
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 456)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T., Ishii,Y. and Hayashizaki,Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .456
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A330084M14"
/sex="male"
/tissue_type="spinal cord"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult male spinal cord"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. "

ORIGIN

Alignment Scores:

Pred. No.:	1.48e-31	Length:	456
Score:	305.00	Matches:	56
Percent Similarity:	91.80%	Conservative:	0
Best Local Similarity:	91.80%	Mismatches:	5
Query Match:	93.56%	Indels:	0
DB:	2	Gaps:	0

US-09-774-954-3 (1-61) x BB633858 (1-456)

```
Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   ||||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db         91 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 150

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        151 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 210

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 270

Qy         61 His 61
              |||
Db        271 CAT 273
```

RESULT 5
BB851550
LOCUS BB851550 480 bp mRNA linear EST 26-NOV-2001
DEFINITION BB851550 RIKEN full-length enriched, B16 F10Y cells Mus musculus cDNA clone G370002F02 5', mRNA sequence.
ACCESSION BB851550
VERSION BB851550.1 GI:17093004
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 480)
AUTHORS Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,

Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .480
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="G370002F02"
/cell_type="B16 F10Y cells"
/clone_lib="RIKEN full-length enriched, B16 F10Y cells"

ORIGIN

Alignment Scores:

Pred. No.:	1.58e-31	Length:	480
Score:	305.00	Matches:	56
Percent Similarity:	91.80%	Conservative:	0
Best Local Similarity:	91.80%	Mismatches:	5
Query Match:	93.56%	Indels:	0
DB:	2	Gaps:	0

US-09-774-954-3 (1-61) x BB851550 (1-480)

```
Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   ||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db        117 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 176

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        177 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 236

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 296

Qy         61 His 61
              |||
Db        297 CAT 299
```

RESULT 6
CF155036

```
LOCUS       CF155036                 536 bp    mRNA    linear   EST 25-JUL-2003
DEFINITION  B0620C09-5 NIA Mouse Blastocyst cDNA Library (Long) Mus musculus
            cDNA clone NIA:B0620C09 IMAGE:30450176 5', mRNA sequence.
ACCESSION   CF155036
VERSION     CF155036.1  GI:33264585
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 536)
  AUTHORS   Piao,Y., Ko,N.T., Lim,M.K. and Ko,M.S.H.
  TITLE     Construction of long-transcript enriched cDNA libraries from
            submicrogram amounts of total RNAs by a universal PCR amplification
            method
  JOURNAL   Genome Res. 11 (9), 1553-1558 (2001)
  MEDLINE   21429098
   PUBMED   11544199
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: B0620  row: C  column: 09
            Seq primer: M13 Reverse
            High quality sequence stop: 536
            POLYA=No.
FEATURES    Location/Qualifiers
     source 1. .536
            /organism="Mus musculus"
            /mol_type="mRNA"
            /strain="C57BL/6J"
            /db_xref="niaEST:B0620C09-5"
            /db_xref="taxon:10090"
            /clone="NIA:B0620C09 IMAGE:30450176"
            /tissue_type="Blastocyst"
            /dev_stage="3.5-dpc"
            /lab_host="DH10B"
            /clone_lib="NIA Mouse Blastocyst cDNA Library (Long)"
            /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were extracted from a pool of 20 Blastocysts. Double-stranded cDNAs were synthesized with an Oligo(dT) primer [Invitrogen: 5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT-3'] from 0.2 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lone-linker LL-Sal4, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.2 kb. The library was constructed by Yulan Piao (NIA)."

ORIGIN
```

Alignment Scores:

Pred. No.:	1.8e-31	Length:	536
Score:	305.00	Matches:	56
Percent Similarity:	91.80%	Conservative:	0
Best Local Similarity:	91.80%	Mismatches:	5
Query Match:	93.56%	Indels:	0
DB:	7	Gaps:	0

US-09-774-954-3 (1-61) x CF155036 (1-536)

```
Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   ||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db         84 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 143

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        144 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 203

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 263

Qy         61 His 61
              |||
Db        264 CAT 266
```

RESULT 7
BB618795

```
LOCUS       BB618795                 604 bp    mRNA    linear   EST 31-AUG-2001
DEFINITION  BB618795 RIKEN full-length enriched, 8 days embryo Mus musculus
```

cDNA clone 5730432G02 5', mRNA sequence.
ACCESSION BB618795
VERSION BB618795.1 GI:15396691
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 604)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T., Ishii,Y. and Hayashizaki,Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.
FEATURES Location/Qualifiers
source 1. .604
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5730432G02"
/sex="mixed"
/dev_stage="8 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 8 days embryo"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to
Rot = 100.0 Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."
ORIGIN

Alignment Scores:

Pred. No.:	2.09e-31	Length:	604
Score:	305.00	Matches:	56
Percent Similarity:	91.80%	Conservative:	0
Best Local Similarity:	91.80%	Mismatches:	5
Query Match:	93.56%	Indels:	0
DB:	2	Gaps:	0

US-09-774-954-3 (1-61) x BB618795 (1-604)

```
Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   |||||||||||||||||||||||||||||||||||||   |||   ||||||||
Db        110 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 169

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        170 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 229

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 289

Qy         61 His 61
              |||
Db        290 CAT 292
```

RESULT 8
BY728018
LOCUS BY728018 605 bp mRNA linear EST 17-DEC-2002
DEFINITION BY728018 RIKEN full-length enriched, 10 days neonate cerebellum Mus musculus cDNA clone B930076G19 5', mRNA sequence.
ACCESSION BY728018
VERSION BY728018.1 GI:27141145
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 605)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,

Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
FEATURES Location/Qualifiers
source 1. .605
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="B930076G19"
/sex="mixed"
/tissue_type="cerebellum"
/dev_stage="10 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 10 days neonate cerebellum"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"
ORIGIN

Alignment Scores:

Pred. No.:	2.09e-31	Length:	605
Score:	305.00	Matches:	56
Percent Similarity:	91.80%	Conservative:	0
Best Local Similarity:	91.80%	Mismatches:	5
Query Match:	93.56%	Indels:	0
DB:	6	Gaps:	0

US-09-774-954-3 (1-61) x BY728018 (1-605)

```
Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   ||||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db         91 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 150

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        151 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 210

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 270

Qy         61 His 61
              |||
Db        271 CAT 273
```

RESULT 9
BB623905
LOCUS BB623905 611 bp mRNA linear EST 26-OCT-2001
DEFINITION BB623905 RIKEN full-length enriched, adult male adrenal gland Mus musculus cDNA clone 7330430I07 5', mRNA sequence.
ACCESSION BB623905
VERSION BB623905.1 GI:16462515
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 611)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)

```
           Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
           Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
           Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
           Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
           further details.
            e mouse tissues.
FEATURES             Location/Qualifiers
     source          1. .611
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="7330430I07"
                     /sex="male"
                     /tissue_type="adrenal gland"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male adrenal
                     gland"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 185.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cloned into the XhoI and BamHI sites.
                     Vector: a modified pBluescript KS(+) after bulk excision
                     from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end:
                     BamHI"
ORIGIN

Alignment Scores:
Pred. No.:                2.12e-31        Length:         611
Score:                    305.00          Matches:        56
Percent Similarity:       91.80%          Conservative:   0
Best Local Similarity:    91.80%          Mismatches:     5
Query Match:              93.56%          Indels:         0
DB:                       2               Gaps:           0

US-09-774-954-3 (1-61) x BB623905 (1-611)

Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   |||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db        117 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 176

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        177 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 236

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 296

Qy         61 His 61
              |||
Db        297 CAT 299

RESULT 10
BY752500
LOCUS       BY752500                  614 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY752500 RIKEN full-length enriched, 9.5 days embryo parthenogenote
            Mus musculus cDNA clone B130006A17 5', mRNA sequence.
ACCESSION   BY752500
VERSION     BY752500.1  GI:27183404
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 614)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I.,
            Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A.,
            Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C.,
            Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H.,
            Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V.,
            Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A.,
            Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T.,
            Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S.,
            Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A.,
            Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A.,
            Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R.,
            Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T.,
            Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G.,
            Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S.,
            Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M.,
            Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K.,
            Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M.,
            Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y.,
            Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I.,
            Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P.,
            Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M.,
            Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.,
            Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y.,
            Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K.,
            Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S.,
            Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P.,
            Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F.,
            Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y.,
            Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M.,
            Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N.,
            Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M.,
            Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
            Direct Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Tissues were provided by Dr. Tomohiro Kono (Department of Animal
            Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
```

Kanagawa Prefecture, Japan) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .614
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="B130006A17"
/tissue_type="parthenogenote"
/dev_stage="9.5 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 9.5 days embryo parthenogenote"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 229.0 Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Alignment Scores:

Pred. No.:	2.13e-31	Length:	614
Score:	305.00	Matches:	56
Percent Similarity:	91.80%	Conservative:	0
Best Local Similarity:	91.80%	Mismatches:	5
Query Match:	93.56%	Indels:	0
DB:	6	Gaps:	0

US-09-774-954-3 (1-61) x BY752500 (1-614)

```
Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   |||||||||||||||||||||||||||||||||||||||   |||   ||||||||
Db        117 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 176

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        177 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 236

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 296

Qy         61 His 61
              |||
Db        297 CAT 299
```

RESULT 11
BB631131
LOCUS BB631131 634 bp mRNA linear EST 26-OCT-2001
DEFINITION BB631131 RIKEN full-length enriched, 16 days neonate thymus Mus musculus cDNA clone A130056K10 5', mRNA sequence.
ACCESSION BB631131
VERSION BB631131.1 GI:16467953
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 634)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .634
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A130056K10"
/tissue_type="thymus"
/dev_stage="16 days neonate"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 16 days neonate thymus"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Alignment Scores:

Pred. No.:	2.21e-31	Length:	634

```
Score:                 305.00        Matches:       56
Percent Similarity:    91.80%        Conservative:  0
Best Local Similarity: 91.80%        Mismatches:    5
Query Match:           93.56%        Indels:        0
DB:                    2             Gaps:          0

US-09-774-954-3 (1-61) x BB631131 (1-634)

Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   |||||||||||||||||||||||||||||||||||   |||   |||||||||
Db        104 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 163

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        164 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 223

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        224 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 283

Qy         61 His 61
              |||
Db        284 CAT 286

RESULT 12
BB611933
LOCUS       BB611933                 641 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB611933 RIKEN full-length enriched, 15 days embryo head Mus
            musculus cDNA clone 4021401D01 5', mRNA sequence.
ACCESSION   BB611933
VERSION     BB611933.1  GI:16453020
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
            Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A.
            and Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
            Sugahara,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I.,
            Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES              Location/Qualifiers
     source           1. .641
                      /organism="Mus musculus"
                      /mol_type="mRNA"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="4021401D01"
                      /sex="mixed"
                      /tissue_type="head"
                      /dev_stage="15 days embryo"
                      /lab_host="DH10B"
                      /clone_lib="RIKEN full-length enriched, 15 days embryo
                      head"
                      /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                      prepared and sequenced in Mouse Genome Encyclopedia
                      Project of Genome Exploration Research Group in Riken
                      Genomic Sciences Center and Genome Science Laboratory in
                      RIKEN. Division of Experimental Animal Research in Riken
                      contributed to prepare mouse tissues. 1st strand cDNA was
                      primed with a primer [5'
                      GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                      prepared by using trehalose thermo-activated reverse
                      transcriptase and subsequently enriched for full-length by
                      cap-trapper. Second strand cDNA was prepared with the
                      primer adapter of sequence [5'
                      GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA
                      was cloned into the XhoI and BamHI sites. Vector: a
                      modified pBluescript KS(+) after bulk excision from Lambda
                      FLC I"
ORIGIN

Alignment Scores:
Pred. No.:             2.24e-31      Length:        641
Score:                 305.00        Matches:       56
Percent Similarity:    91.80%        Conservative:  0
Best Local Similarity: 91.80%        Mismatches:    5
Query Match:           93.56%        Indels:        0
DB:                    2             Gaps:          0

US-09-774-954-3 (1-61) x BB611933 (1-641)

Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   |||||||||||||||||||||||||||||||||||   |||   |||||||||
Db         91 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 150

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        151 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 210

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 270

Qy         61 His 61
              |||
Db        271 CAT 273

RESULT 13
BY752497
LOCUS       BY752497                 650 bp    mRNA    linear   EST 17-DEC-2002
DEFINITION  BY752497 RIKEN full-length enriched, 9.5 days embryo parthenogenote
            Mus musculus cDNA clone B130001I18 5', mRNA sequence.
ACCESSION   BY752497
VERSION     BY752497.1  GI:27183400
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 650)

AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

MEDLINE 22354683

PUBMED 12466851

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Tissues were provided by Dr. Tomohiro Kono (Department of Animal Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City, Kanagawa Prefecture, Japan) whose assistance we gratefully acknowledge.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers

source 1. .650
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="B130001I18"
/tissue_type="parthenogenote"
/dev_stage="9.5 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 9.5 days embryo parthenogenote"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 229.0 Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

ORIGIN

Alignment Scores:

Pred. No.:	2.28e-31	Length:	650
Score:	305.00	Matches:	56
Percent Similarity:	91.80%	Conservative:	0
Best Local Similarity:	91.80%	Mismatches:	5
Query Match:	93.56%	Indels:	0
DB:	6	Gaps:	0

US-09-774-954-3 (1-61) x BY752497 (1-650)

```
Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||  ||||||||||||||||||||||||||||||||||||||||  |||  |||||||||
Db        117 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 176

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        177 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 236

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        237 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 296

Qy         61 His 61
              |||
Db        297 CAT 299
```

RESULT 14
BB642654

```
LOCUS       BB642654                661 bp    mRNA    linear   EST 31-AUG-2001
DEFINITION  BB642654 RIKEN full-length enriched, adult retina Mus musculus cDNA
            clone A930028F21 5', mRNA sequence.
ACCESSION   BB642654
VERSION     BB642654.1  GI:15401970
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T.,
            Hara,A., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J.,
            Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K.,
            Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
```

Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. . 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. . 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
Yamanaka,I., Kiyosawa,H., Kondo,S., Saito,T., Shinagawa,A., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K., Arakawa,T., Ishii,Y. and Hayashizaki,Y.
Mapping of 19032 mouse cDNAs on mouse chromosomes. J. Struct. Func. Genomics 2 pre, L72-L86 (2001
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.
e mouse tissues.

FEATURES Location/Qualifiers
source 1. .661
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="A930028F21"
/tissue_type="retina"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, adult retina"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5'GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. ~Retina RNA was provided by Stefano Gustincich, Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA02115, USA, whose assistance we gratefully acknowledge."

ORIGIN

Alignment Scores:

Pred. No.:	2.33e-31	Length:	661
Score:	305.00	Matches:	56
Percent Similarity:	91.80%	Conservative:	0
Best Local Similarity:	91.80%	Mismatches:	5
Query Match:	93.56%	Indels:	0
DB:	2	Gaps:	0

US-09-774-954-3 (1-61) x BB642654 (1-661)

```
Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   ||||||||||||||||||||||||||||||||||||    |||   |||||||||
Db        118 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 177

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        178 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 237

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        238 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 297

Qy         61 His 61
              |||
Db        298 CAT 300
```

RESULT 15
BY728337
LOCUS BY728337 664 bp mRNA linear EST 17-DEC-2002
DEFINITION BY728337 RIKEN full-length enriched, 16 days embryo head Mus musculus cDNA clone C130003G23 5', mRNA sequence.
ACCESSION BY728337
VERSION BY728337.1 GI:27141464
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 664)
AUTHORS Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H., Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C., Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A., Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S., Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A., Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J., Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M., King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G., Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D., Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y., Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A., Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K., Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 22354683
PUBMED 12466851
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222

Fax: 81-45-503-9216
Email: genome-res@gsc.riken.jp, URL:http://genome.gsc.riken.jp/
Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

FEATURES Location/Qualifiers
source 1. .664
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="C130003G23"
/sex="mixed"
/tissue_type="head"
/dev_stage="16 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 16 days embryo head"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I"

ORIGIN

Alignment Scores:

Pred. No.:	2.34e-31	Length:	664
Score:	305.00	Matches:	56
Percent Similarity:	91.80%	Conservative:	0
Best Local Similarity:	91.80%	Mismatches:	5
Query Match:	93.56%	Indels:	0
DB:	6	Gaps:	0

US-09-774-954-3 (1-61) x BY728337 (1-664)

```
Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   ||||||||||||||||||||||||||||||||||||   |||   |||||||||
Db         91 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 150

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db        151 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 210

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 270

Qy         61 His 61
              |||
Db        271 CAT 273
```

Search completed: October 26, 2005, 15:17:23
Job time : 1076.43 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 09:31:14 ; Search time 260.659 Seconds
(without alignments)
1931.299 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGYLLYXPXMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9772377 seqs, 4126317084 residues

Total number of hits satisfying chosen parameters: 19544754

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105434_6433/app_query.fasta_1.917
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09774954_@CGN_1_1_684_@runat_25102005_105434_6433
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	295	90.5	1100	11	US-09-774-954-16	Sequence 16, Appl
2	295	90.5	1300	11	US-09-774-954-4	Sequence 4, Appli
3	295	90.5	1400	22	US-10-956-157-9986	Sequence 9986, Ap
4	295	90.5	1506	22	US-10-956-157-4751	Sequence 4751, Ap
5	295	90.5	1514	11	US-09-774-954-1	Sequence 1, Appli
6	295	90.5	5218	16	US-10-301-822-168	Sequence 168, App
7	295	90.5	5218	22	US-10-956-157-4752	Sequence 4752, Ap
8	295	90.5	11284	11	US-09-774-954-5	Sequence 5, Appli
9	204	62.6	5009	11	US-09-774-954-7	Sequence 7, Appli
10	177	54.3	1209	26	US-11-097-143-18161	Sequence 18161, A
c 11	145	44.5	3264	26	US-11-097-143-18160	Sequence 18160, A
c 12	145	44.5	3793	26	US-11-097-143-18082	Sequence 18082, A
c 13	76	23.3	521	13	US-09-925-065A-602444	Sequence 602444,
14	76	23.3	760	24	US-10-820-474A-236	Sequence 236, App
15	72.5	22.2	1803	16	US-10-156-761-5001	Sequence 5001, Ap
16	72.5	22.2	9025608	16	US-10-156-761-1	Sequence 1, Appli
17	71.5	21.9	568	21	US-10-425-115-16666	Sequence 16666, A
18	69.5	21.3	594	18	US-10-282-122A-25606	Sequence 25606, A
c 19	69	21.2	764	21	US-10-425-115-28075	Sequence 28075, A
20	68.5	21.0	829	19	US-10-424-599-123276	Sequence 123276,
c 21	68	20.9	4953	20	US-10-437-963-23967	Sequence 23967, A
22	67.5	20.7	1617	20	US-10-437-963-11436	Sequence 11436, A
c 23	66.5	20.4	636	13	US-09-925-065A-285729	Sequence 285729,
24	66.5	20.4	2223	22	US-10-472-928-1697	Sequence 1697, Ap
25	66.5	20.4	2292	24	US-10-617-320-1007	Sequence 1007, Ap
26	66.5	20.4	2304	9	US-09-815-242-9245	Sequence 9245, Ap
27	66.5	20.4	2304	9	US-09-815-242-9500	Sequence 9500, Ap
28	66.5	20.4	2304	10	US-09-769-787-308	Sequence 308, App
29	66.5	20.4	2304	18	US-10-282-122A-37727	Sequence 37727, A
30	66.5	20.4	2352	9	US-09-775-978-1	Sequence 1, Appli
31	66.5	20.4	18627	8	US-08-961-527-113	Sequence 113, App
32	66.5	20.4	18627	18	US-10-158-844-113	Sequence 113, App
33	66.5	20.4	2162598	22	US-10-472-928-4979	Sequence 4979, Ap
34	66	20.2	361	13	US-09-925-065A-476650	Sequence 476650,
c 35	65	19.9	478	19	US-10-152-319A-21	Sequence 21, Appl
c 36	65	19.9	519	20	US-10-437-963-80416	Sequence 80416, A
37	65	19.9	1437	19	US-10-424-599-40441	Sequence 40441, A
38	65	19.9	2910	19	US-10-424-599-40437	Sequence 40437, A
39	65	19.9	513509	10	US-09-754-853A-4	Sequence 4, Appli
40	64.5	19.8	476	13	US-09-925-065A-500787	Sequence 500787,
41	64.5	19.8	476	13	US-09-925-065A-500788	Sequence 500788,
c 42	64.5	19.8	575	24	US-10-631-467-1068	Sequence 1068, Ap
c 43	64.5	19.8	628	13	US-09-925-065A-770388	Sequence 770388,
c 44	64	19.6	600	24	US-10-972-079-32895	Sequence 32895, A
45	64	19.6	1018	18	US-10-282-122A-36872	Sequence 36872, A

ALIGNMENTS

RESULT 1
US-09-774-954-16
; Sequence 16, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA

```
;           ZIP: 94080
;        COMPUTER READABLE FORM:
;           MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/09/774,954
;           FILING DATE: 30-Jan-2001
;           CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US/08/978,741
;           FILING DATE: 26-NOV-1997
;           APPLICATION NUMBER: 08/792,498
;           FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;           NAME: Svoboda, Craig G.
;           REGISTRATION NUMBER: 39,044
;           REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;           TELEPHONE: 650/225-1489
;           TELEFAX: 650/952-9881
;     INFORMATION FOR SEQ ID NO: 16:
;        SEQUENCE CHARACTERISTICS:
;           LENGTH: 1100 base pairs
;           TYPE: Nucleic Acid
;           STRANDEDNESS: Single
;           TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-774-954-16

Alignment Scores:
Pred. No.:              3.65e-33        Length:         1100
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     11              Gaps:           0

US-09-774-954-3 (1-61) x US-09-774-954-16 (1-1100)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              |||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db          7 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 66

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              |||||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db         67 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 126

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 183

RESULT 2
US-09-774-954-4
; Sequence 4, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;           ADDRESSEE: Genentech, Inc.
;           STREET: 1 DNA Way
;           CITY: South San Francisco
;           STATE: California
;           COUNTRY: USA
;           ZIP: 94080
;        COMPUTER READABLE FORM:
;           MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/09/774,954
;           FILING DATE: 30-Jan-2001
;           CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: US/08/978,741
;           FILING DATE: 26-NOV-1997
;           APPLICATION NUMBER: 08/792,498
;           FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;           NAME: Svoboda, Craig G.
;           REGISTRATION NUMBER: 39,044
;           REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;           TELEPHONE: 650/225-1489
;           TELEFAX: 650/952-9881
;     INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;           LENGTH: 1300 base pairs
;           TYPE: Nucleic Acid
;           STRANDEDNESS: Single
;           TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-774-954-4

Alignment Scores:
Pred. No.:              4.42e-33        Length:         1300
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     11              Gaps:           0

US-09-774-954-3 (1-61) x US-09-774-954-4 (1-1300)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              |||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db        142 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 201

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              |||||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        202 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 261

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 318

RESULT 3
US-10-956-157-9986
; Sequence 9986, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William
;  TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;  TITLE OF INVENTION:  HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;  FILE REFERENCE: 031896-043000 (AM 101081)
;  CURRENT APPLICATION NUMBER: US/10/956,157
;  CURRENT FILING DATE:  2004-10-04
;  NUMBER OF SEQ ID NOS: 319805
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9986
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-9986

Alignment Scores:
Pred. No.:              4.82e-33        Length:         1400
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
```

```
DB:                    22            Gaps:          0

US-09-774-954-3 (1-61) x US-10-956-157-9986 (1-1400)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||  ||||||||||||||||||  |||  ||||||||||||||||
Db         19 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 78

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db         79 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 138

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 195

RESULT 4
US-10-956-157-4751
; Sequence 4751, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William
;  TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;  TITLE OF INVENTION:  HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;  FILE REFERENCE: 031896-043000 (AM 101081)
;  CURRENT APPLICATION NUMBER: US/10/956,157
;  CURRENT FILING DATE:  2004-10-04
;  NUMBER OF SEQ ID NOS: 319805
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4751
;   LENGTH: 1506
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-4751

Alignment Scores:
Pred. No.:              5.24e-33      Length:          1506
Score:                  295.00        Matches:         54
Percent Similarity:     91.53%        Conservative:    0
Best Local Similarity:  91.53%        Mismatches:      5
Query Match:            90.49%        Indels:          0
DB:                     22            Gaps:            0

US-09-774-954-3 (1-61) x US-10-956-157-4751 (1-1506)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||  ||||||||||||||||||  |||  ||||||||||||||||
Db        125 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 184

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        185 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 244

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 301

RESULT 5
US-09-774-954-1
; Sequence 1, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1514 base pairs
;             TYPE: Nucleic Acid
;             STRANDEDNESS: Single
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-774-954-1

Alignment Scores:
Pred. No.:              5.27e-33      Length:          1514
Score:                  295.00        Matches:         54
Percent Similarity:     91.53%        Conservative:    0
Best Local Similarity:  91.53%        Mismatches:      5
Query Match:            90.49%        Indels:          0
DB:                     11            Gaps:            0

US-09-774-954-3 (1-61) x US-09-774-954-1 (1-1514)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||  ||||||||||||||||||  |||  ||||||||||||||||
Db          7 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 66

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db         67 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 126

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 183

RESULT 6
US-10-301-822-168
; Sequence 168, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
;  APPLICANT:  Burgart, Lawrence J.
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
;  CURRENT APPLICATION NUMBER: US/10/301,822
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
```

```
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 168
;   LENGTH: 5218
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (50)...(1216)
US-10-301-822-168

Alignment Scores:
Pred. No.:              2.18e-32        Length:         5218
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     16              Gaps:           0

US-09-774-954-3 (1-61) x US-10-301-822-168 (1-5218)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              |||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db        125 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 184

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              |||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        185 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 244

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 301

RESULT 7
US-10-956-157-4752
; Sequence 4752, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William
;  TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;  TITLE OF INVENTION:  HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;  FILE REFERENCE: 031896-043000 (AM 101081)
;  CURRENT APPLICATION NUMBER: US/10/956,157
;  CURRENT FILING DATE:  2004-10-04
;  NUMBER OF SEQ ID NOS: 319805
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4752
;   LENGTH: 5218
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-4752

Alignment Scores:
Pred. No.:              2.18e-32        Length:         5218
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     22              Gaps:           0

US-09-774-954-3 (1-61) x US-10-956-157-4752 (1-5218)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              |||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db        125 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 184

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              |||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        185 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 244

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 301

RESULT 8
US-09-774-954-5
; Sequence 5, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 5:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 11284 base pairs
;             TYPE: Nucleic Acid
;             STRANDEDNESS: Single
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-774-954-5

Alignment Scores:
Pred. No.:              5.29e-32        Length:         11284
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     11              Gaps:           0

US-09-774-954-3 (1-61) x US-09-774-954-5 (1-11284)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              |||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db       4242 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 4301

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              |||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db       4302 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 4361

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4362 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 4418

Db        185 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 244

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 301
```

```
RESULT 9
US-09-774-954-7
; Sequence 7, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 7:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 5009 base pairs
;             TYPE: Nucleic Acid
;             STRANDEDNESS: Single
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-774-954-7

Alignment Scores:
Pred. No.:              5.28e-19        Length:         5009
Score:                  204.00          Matches:        37
Percent Similarity:     94.87%          Conservative:   0
Best Local Similarity:  94.87%          Mismatches:     2
Query Match:            62.58%          Indels:         0
DB:                     11              Gaps:           0

US-09-774-954-3 (1-61) x US-09-774-954-7 (1-5009)

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db          2 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 61

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 118

RESULT 10
US-11-097-143-18161
; Sequence 18161, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18161
;   LENGTH: 1209
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-18161

Alignment Scores:
Pred. No.:              9.79e-16        Length:         1209
Score:                  177.00          Matches:        33
Percent Similarity:     80.00%          Conservative:   3
Best Local Similarity:  73.33%          Mismatches:     9
Query Match:            54.29%          Indels:         0
DB:                     26              Gaps:           0

US-09-774-954-3 (1-61) x US-11-097-143-18161 (1-1209)

Qy          7 AspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyAsnGlnAlaAsp 26
              |||       |||||||||   |||   |||   ||||||||||||||||||||||||||
Db         82 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCGAC 141

Qy         27 HisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProPro 46
              ||||||||||||||||||||||||||||||        |||||||||   :::||||||
Db        142 CACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCG 201

Qy         47 TrpIleGluTyrGln 51
              |||:::||||||:::
Db        202 TGGGTGGAGTATCGT 216

RESULT 11
US-11-097-143-18160/c
; Sequence 18160, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
```

```
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18160
;   LENGTH: 3264
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-18160

Alignment Scores:
Pred. No.:              1.58e-10        Length:         3264
Score:                  145.00          Matches:        26
Percent Similarity:     87.88%          Conservative:   3
Best Local Similarity:  78.79%          Mismatches:     4
Query Match:            44.48%          Indels:         0
DB:                     26              Gaps:           0

US-09-774-954-3 (1-61) x US-11-097-143-18160 (1-3264)

Qy         19 GlyArgPheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu*** 38
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2092 GGACGCTTTGGCAACCAGGCCGACCACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTT 2033

Qy         39 ValArgThrLeuAlaValProProTrpIleGluTyrGln 51
                |||||||||   :::|||||||||:::||||||:::
Db       2032 AATCGCACCCTGATCCTGCCGCCGTGGGTGGAGTATCGT 1994

RESULT 12
US-11-097-143-18082/c
; Sequence 18082, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18082
;   LENGTH: 3793
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-18082

Alignment Scores:
Pred. No.:              1.88e-10        Length:         3793
Score:                  145.00          Matches:        26
Percent Similarity:     87.88%          Conservative:   3
Best Local Similarity:  78.79%          Mismatches:     4
Query Match:            44.48%          Indels:         0
DB:                     26              Gaps:           0

US-09-774-954-3 (1-61) x US-11-097-143-18082 (1-3793)

Qy         19 GlyArgPheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu*** 38
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        584 GGACGCTTTGGCAACCAGGCCGACCACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTT 525

Qy         39 ValArgThrLeuAlaValProProTrpIleGluTyrGln 51
                |||||||||   :::|||||||||:::||||||:::
Db        524 AATCGCACCCTGATCCTGCCGCCGTGGGTGGAGTATCGT 486

RESULT 13
US-09-925-065A-602444/c
; Sequence 602444, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 602444
;   LENGTH: 521
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-602444

Alignment Scores:
Pred. No.:              0.281           Length:         521
Score:                  76.00           Matches:        14
Percent Similarity:     83.33%          Conservative:   1
Best Local Similarity:  77.78%          Mismatches:     3
Query Match:            23.31%          Indels:         0
DB:                     13              Gaps:           0

US-09-774-954-3 (1-61) x US-09-925-065A-602444 (1-521)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||||||||||||||   ||||||||||||||||||   |||   ||||||:::
Db        280 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGTAAG 227

RESULT 14
US-10-820-474A-236
; Sequence 236, Application US/10820474A
; Publication No. US20050155089A1
; GENERAL INFORMATION:
;  APPLICANT: LAL, PREETI
;  APPLICANT:  TANG, Y. TOM
;  APPLICANT:  GORGONE, GINA A.
;  APPLICANT:  CORLEY, NEIL C.
;  APPLICANT:  GUEGLER, KARL J.
;  APPLICANT:  BAUGHN, MARIAH R.
;  APPLICANT:  AKERBLOM, INGRID E.
;  APPLICANT:  AU-YOUNG, JANICE
;  APPLICANT:  YUE, HENRY
;  APPLICANT:  PATTERSON, CHANDRA
;  APPLICANT:  REDDY, ROOPA
;  APPLICANT:  HILLMAN, JENNIFER L.
```

```
;  APPLICANT:  BANDMAN, OLGA
;  TITLE OF INVENTION: SIGNAL PEPTIDE-CONTAINING MOLECULES
;  FILE REFERENCE: 039386-1568
;  CURRENT APPLICATION NUMBER: US/10/820,474A
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: 09/720,533
;  PRIOR FILING DATE: 2001-03-20
;  PRIOR APPLICATION NUMBER: PCT/US99/14484
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/090,762
;  PRIOR FILING DATE: 1998-06-26
;  PRIOR APPLICATION NUMBER: 60/094,983
;  PRIOR FILING DATE: 1998-07-31
;  PRIOR APPLICATION NUMBER: 60/102,686
;  PRIOR FILING DATE: 1998-10-01
;  NUMBER OF SEQ ID NOS: 269
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 236
;   LENGTH: 760
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte Clone No: 1647884
US-10-820-474A-236

Alignment Scores:
Pred. No.:              0.433          Length:         760
Score:                  76.00          Matches:        14
Percent Similarity:     83.33%         Conservative:   1
Best Local Similarity:  77.78%         Mismatches:     3
Query Match:            23.31%         Indels:         0
DB:                     24             Gaps:           0

US-09-774-954-3 (1-61) x US-10-820-474A-236 (1-760)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              ||||||||||||||   ||||||||||||||||||   |||   ||||||:::
Db        117 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGTAAG 170

RESULT 15
US-10-156-761-5001
; Sequence 5001, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5001
;   LENGTH: 1803
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1803)
US-10-156-761-5001

Alignment Scores:
Pred. No.:              3.83           Length:         1803
Score:                  72.50          Matches:        23
Percent Similarity:     40.26%         Conservative:   8
Best Local Similarity:  29.87%         Mismatches:     17
Query Match:            22.24%         Indels:         29
DB:                     16             Gaps:           5

US-09-774-954-3 (1-61) x US-10-156-761-5001 (1-1803)

Qy          5 SerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyAsnGln 24
              |||||||||:::      |||   |||:::   |||   :::|||   :::|||   |||
Db        178 TCCTGGGACATCAAGAACTAT---CTGTGGGCGCCCAGGCTCGGCATGTACGGCATCCAG 234

Qy         25 AlaAspHisPheLeu-------------------------------------GlySer 31
                 |||   |||                                        ||||||
Db        235 CGCATCCACCTCCTGGGCAACGTCATGGTGCTGGCGGGCGCCGGGGTCGGCGGCGGCTCC 294

Qy         32 LeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProPro--------------- 46
              |||   :::|||              ||||||   |||||||||
Db        295 CTCAACTACGCG------------AACACCCTCTACGTACCGCCGAAGGCCTTCTTCGAC 342

Qy         47 ---------Trp---------IleGluTyrGlnHisHisLysProProPhe 57
                      |||     .     ::::::|||           |||:::
Db        343 GACCCGCAGTGGAAGGACATCACCGACTGGCAGGAGGAACTGAAGCCGTAC 393

Search completed: October 26, 2005, 16:00:04
Job time : 287.326 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 08:18:39 ; Search time 53.4846 Seconds
(without alignments)
1866.199 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGYLLYXPXMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105433_6415/app_query.fasta_1.917
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954_@CGN_1_1_116_@runat_25102005_105433_6415 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	295	90.5	1100	3	US-08-978-741-16	Sequence 16, Appl
2	295	90.5	1100	3	US-09-333-729A-16	Sequence 16, Appl
3	295	90.5	1300	3	US-08-978-741-4	Sequence 4, Appli
4	295	90.5	1300	3	US-09-333-729A-6	Sequence 6, Appli
5	295	90.5	1514	3	US-08-978-741-1	Sequence 1, Appli
6	295	90.5	1514	3	US-09-333-729A-2	Sequence 2, Appli
7	295	90.5	11284	3	US-08-978-741-5	Sequence 5, Appli
8	204	62.6	5009	3	US-08-978-741-7	Sequence 7, Appli
9	204	62.6	5009	3	US-09-333-729A-8	Sequence 8, Appli
10	177	54.3	1320	4	US-09-270-767-14353	Sequence 14353, A
c 11	76	23.3	19227	4	US-09-949-016-12127	Sequence 12127, A
c 12	76	23.3	19228	4	US-09-949-016-16285	Sequence 16285, A
c 13	68.5	21.0	42325	4	US-08-311-731A-131	Sequence 131, App
14	67	20.6	266293	4	US-09-949-016-11934	Sequence 11934, A
15	66.5	20.4	2214	4	US-09-583-110-2414	Sequence 2414, Ap
16	66.5	20.4	2292	4	US-09-107-433-1007	Sequence 1007, Ap
17	66.5	20.4	2352	2	US-08-922-837-1	Sequence 1, Appli
18	66.5	20.4	2352	3	US-09-351-550-1	Sequence 1, Appli
19	66.5	20.4	18627	3	US-08-961-527-113	Sequence 113, App
c 20	65	19.9	601	4	US-09-949-016-203950	Sequence 203950,
c 21	65	19.9	601	4	US-09-949-016-203951	Sequence 203951,
22	65	19.9	276237	4	US-09-949-016-17504	Sequence 17504, A
c 23	64.5	19.8	175265	4	US-09-949-016-16089	Sequence 16089, A
c 24	62.5	19.2	154600	4	US-09-949-016-14757	Sequence 14757, A
c 25	62	19.0	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 26	62	19.0	4411529	3	US-09-103-840A-1	Sequence 1, Appli
27	61.5	18.9	999	5	PCT-US91-00899-10	Sequence 10, Appl
28	61.5	18.9	1136	1	US-08-395-800A-9	Sequence 9, Appli
29	61.5	18.9	1155	1	US-08-434-151-1	Sequence 1, Appli
30	61.5	18.9	1155	1	US-08-208-889A-1	Sequence 1, Appli
31	61.5	18.9	1155	2	US-08-433-271-1	Sequence 1, Appli
32	61.5	18.9	1155	2	US-08-715-259-1	Sequence 1, Appli
33	61.5	18.9	1174	5	PCT-US95-07554-3	Sequence 3, Appli
34	61.5	18.9	1199	1	US-08-395-800A-5	Sequence 5, Appli
35	61.5	18.9	2268	3	US-08-675-773B-4	Sequence 4, Appli
36	61.5	18.9	3373	1	US-08-273-411-2	Sequence 2, Appli
37	61.5	18.9	3374	4	US-09-949-016-5482	Sequence 5482, Ap
38	61.5	18.9	3791	3	US-08-675-773B-3	Sequence 3, Appli
39	61.5	18.9	8174	1	US-07-914-281-5	Sequence 5, Appli
40	61.5	18.9	8174	1	US-08-393-246-5	Sequence 5, Appli
41	61.5	18.9	8174	1	US-08-525-058A-5	Sequence 5, Appli
42	61.5	18.9	8174	2	US-08-696-731-5	Sequence 5, Appli
43	61.5	18.9	8174	3	US-09-042-531-5	Sequence 5, Appli
44	61.5	18.9	8174	5	PCT-US91-00899-3	Sequence 3, Appli
45	61.5	18.9	8580	4	US-09-949-016-17224	Sequence 17224, A

ALIGNMENTS

RESULT 1
US-08-978-741-16
; Sequence 16, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:

```
;      LENGTH:  1100 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-16

Alignment Scores:
Pred. No.:              5.56e-36        Length:         1100
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-3 (1-61) x US-08-978-741-16 (1-1100)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db          7 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 66

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db         67 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 126

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 183

RESULT 2
US-09-333-729A-16
; Sequence 16, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 16
;   LENGTH: 1100
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-16

Alignment Scores:
Pred. No.:              5.56e-36        Length:         1100
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-3 (1-61) x US-09-333-729A-16 (1-1100)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db          7 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 66

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db         67 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 126

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 183

RESULT 3
US-08-978-741-4
; Sequence 4, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1300 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-4

Alignment Scores:
Pred. No.:              6.98e-36        Length:         1300
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-3 (1-61) x US-08-978-741-4 (1-1300)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db        142 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 201

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        202 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 261

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 318

RESULT 4
US-09-333-729A-6
; Sequence 6, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
```

```
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 6
;   LENGTH: 1300
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid insert.
US-09-333-729A-6

Alignment Scores:
Pred. No.:              6.98e-36        Length:         1300
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-3 (1-61) x US-09-333-729A-6 (1-1300)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              |||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db        142 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 201

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||       |||||||||
Db        202 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 261

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 318

RESULT 5
US-08-978-741-1
; Sequence 1, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1514 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-1

Alignment Scores:
Pred. No.:              8.58e-36        Length:         1514
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-3 (1-61) x US-08-978-741-1 (1-1514)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              |||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db          7 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 66

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||       |||||||||
Db         67 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 126

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 183

RESULT 6
US-09-333-729A-2
; Sequence 2, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 2
;   LENGTH: 1514
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-2

Alignment Scores:
Pred. No.:              8.58e-36        Length:         1514
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-3 (1-61) x US-09-333-729A-2 (1-1514)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              |||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db          7 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 66

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||       |||||||||
Db         67 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 126

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 183

RESULT 7
US-08-978-741-5
; Sequence 5, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11284 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-5

Alignment Scores:
Pred. No.:              1.32e-34        Length:         11284
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-3 (1-61) x US-08-978-741-5 (1-11284)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||   |||||||||||||||||   |||   ||||||||||||||||
Db       4242 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 4301

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db       4302 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 4361

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4362 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 4418

RESULT 8
US-08-978-741-7
; Sequence 7, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5009 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-7

Alignment Scores:
Pred. No.:              5.71e-21        Length:         5009
Score:                  204.00          Matches:        37
Percent Similarity:     94.87%          Conservative:   0
Best Local Similarity:  94.87%          Mismatches:     2
Query Match:            62.58%          Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-3 (1-61) x US-08-978-741-7 (1-5009)

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db          2 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 61

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 118

RESULT 9
US-09-333-729A-8
; Sequence 8, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 8
;   LENGTH: 5009
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-8

Alignment Scores:
Pred. No.:              5.71e-21        Length:         5009
Score:                  204.00          Matches:        37
Percent Similarity:     94.87%          Conservative:   0
Best Local Similarity:  94.87%          Mismatches:     2
Query Match:            62.58%          Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-3 (1-61) x US-09-333-729A-8 (1-5009)
```

```
Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db          2 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 61

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 118
```

RESULT 10
US-09-270-767-14353
; Sequence 14353, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14353
; LENGTH: 1320
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-14353

Alignment Scores:			
Pred. No.:	1.43e-17	Length:	1320
Score:	177.00	Matches:	33
Percent Similarity:	80.00%	Conservative:	3
Best Local Similarity:	73.33%	Mismatches:	9
Query Match:	54.29%	Indels:	0
DB:	4	Gaps:	0

US-09-774-954-3 (1-61) x US-09-270-767-14353 (1-1320)

```
Qy          7 AspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyAsnGlnAlaAsp 26
              |||       |||||||||    |||    |||    ||||||||||||||||||||||||
Db         75 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCGAC 134

Qy         27 HisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProPro 46
              ||||||||||||||||||||||||||||||         |||||||||    :::||||||
Db        135 CACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCG 194

Qy         47 TrpIleGluTyrGln 51
              |||:::||||||:::
Db        195 TGGGTGGAGTATCGT 209
```

RESULT 11
US-09-949-016-12127/c
; Sequence 12127, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12127
; LENGTH: 19227
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12127

Alignment Scores:			
Pred. No.:	2.55	Length:	19227
Score:	76.00	Matches:	14
Percent Similarity:	83.33%	Conservative:	1
Best Local Similarity:	77.78%	Mismatches:	3
Query Match:	23.31%	Indels:	0
DB:	4	Gaps:	0

US-09-774-954-3 (1-61) x US-09-949-016-12127 (1-19227)

```
Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              ||||||||||||||||    ||||||||||||||||||    |||    ||||||:::
Db       1712 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGTAAG 1659
```

RESULT 12
US-09-949-016-16285/c
; Sequence 16285, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16285
; LENGTH: 19228
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16285

Alignment Scores:			
Pred. No.:	2.55	Length:	19228
Score:	76.00	Matches:	14
Percent Similarity:	83.33%	Conservative:	1
Best Local Similarity:	77.78%	Mismatches:	3
Query Match:	23.31%	Indels:	0
DB:	4	Gaps:	0

US-09-774-954-3 (1-61) x US-09-949-016-16285 (1-19228)

```
Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              ||||||||||||||||    ||||||||||||||||||    |||    ||||||:::
Db       1712 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGTAAG 1659
```

RESULT 13
US-08-311-731A-131/c
; Sequence 131, Application US/08311731A
; Patent No. 6583266
; GENERAL INFORMATION:
; APPLICANT: SMITH, DOUGLAS
; APPLICANT: MAO, JEN-I
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 411
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/311,731A
;      FILING DATE:
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  GATES, EDWARD R.
;      REGISTRATION NUMBER:  31,616
;      REFERENCE/DOCKET NUMBER:  C0044/7125
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617/720-3500
;      TELEFAX:  617/720-2441
;  INFORMATION FOR SEQ ID NO:  131:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  42325 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  MYCOBACTERIUM LEPRAE
US-08-311-731A-131

Alignment Scores:
Pred. No.:              109          Length:         42325
Score:                  68.50        Matches:        23
Percent Similarity:     42.65%       Conservative:   6
Best Local Similarity:  33.82%       Mismatches:     14
Query Match:            21.01%       Indels:         25
DB:                     4            Gaps:           4

US-09-774-954-3 (1-61) x US-08-311-731A-131 (1-42325)

Qy          5 SerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyAsnGln 24
              |||||||||||||      :::   |||:::   |||   :::|||   :::|||   |||
Db      30298 TCCTGGGATCTGCGTAAATTC---CTCTGGGCGCCGAAGCTGGGCTGCTACGGCATCCAA 30242

Qy         25 AlaAspHisPheLeu----------------------------------------GlySer 31
                 |||   |||                                            ||||||
Db      30241 CGCATTCACCTGCTGCGCAACGTGATGATATTGGCCGGCGCCGGAGTGGGAGGCGGTTCG 30182

Qy         32 LeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProProTrpIleGluTyrGln 51
              |||   :::|||              ||||||   |||||||||
Db      30181 TTGAACTACGCG------------AATACGTTGTACGTTCCGCCG--------------- 30149

Qy         52 HisHisLysProProPheThrAsn 59
                :::|||   |||   |||
Db      30148 ------GAGCCGTTCTTTGCCAAC 30131

RESULT 14
US-09-949-016-11934
; Sequence 11934, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11934
;   LENGTH: 266293
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-11934

Alignment Scores:
Pred. No.:              2.28e+03     Length:         266293
Score:                  67.00        Matches:        19
Percent Similarity:     45.45%       Conservative:   11
Best Local Similarity:  28.79%       Mismatches:     22
Query Match:            20.55%       Indels:         14
DB:                     4            Gaps:           3

US-09-774-954-3 (1-61) x US-09-949-016-11934 (1-266293)

Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||      ||||||   :::   |||:::||||||      |||      ||||||
Db     236166 CGACAACCTGCTTCCTGGTTCATAGATGGCCATCTTCTCACTCTACCTTCACATGGCAGA 236225

Qy         21 PheGlyAsnGlnAlaAspHisPheLeu---------GlySerLeuAlaPheAlaLys--- 36
                |||   :::::   :::          |||      :::   |||:::
Db     236226 AAGGGAGGAGAGAGCTCTCTGTGGTCTCTTTTATGAGGGCTCTACTCTCATGCAAGAGAC 236285

Qy         37 Leu***ValArgThrLeuAlaValProProTrp-------------------------- 47
              |||   :::      |||   :::|||   |||
Db     236286 CTCCACCTTCATGACTTAGTCATCCCCCAGTGGTCCCACCTTCTAATACCATCACCTCAG 236345

Qy         48 ---IleGluTyrGlnHis 52
                 :::|||:::||||||
Db     236346 GAGTTAGAATTTCAACAT 236363

RESULT 15
US-09-583-110-2414
; Sequence 2414, Application US/09583110
; Patent No. 6699703
; GENERAL INFORMATION:
;  APPLICANT:  Lynn Doucette-Stamm et al.
;  TITLE OF INVENTION:  Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
;  TITLE OF INVENTION:  Pneumoniae for Diagnostics and Therapeutics
;  FILE REFERENCE:  PATH00-07A
;  CURRENT APPLICATION NUMBER: US/09/583,110
;  CURRENT FILING DATE:  2000-05-26
;  PRIOR APPLICATION NUMBER:  US 09/107,433
;  PRIOR FILING DATE:  1998-06-30
;  PRIOR APPLICATION NUMBER:  US 60/085,131
;  PRIOR FILING DATE:  1998-05-12
;  PRIOR APPLICATION NUMBER:  US 60/051,553
;  PRIOR FILING DATE:  1997-07-02
;  NUMBER OF SEQ ID NOS:  5322
; SEQ ID NO 2414
;   LENGTH: 2214
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
US-09-583-110-2414

Alignment Scores:
Pred. No.:              4.01         Length:         2214
Score:                  66.50        Matches:        16
Percent Similarity:     48.84%       Conservative:   5
Best Local Similarity:  37.21%       Mismatches:     19
Query Match:            20.40%       Indels:         3
DB:                     4            Gaps:           1

US-09-774-954-3 (1-61) x US-09-583-110-2414 (1-2214)

Qy          8 LeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyAsnGlnAlaAspHis 27
              |||   |||      ||||||   |||          |||   |||        :::
Db        418 TTGATCGGGGTCGCTCTTTATATTCCAACAGCCTTTCTCTTTTCAAATATCGGAACTTAC 477

Qy         28 PheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProProTrp 47
              |||:::||||||:::          :::   |||   :::|||   |||   ||||||
Db        478 TTTATTGGTTCTATCTTG---------ATTTTAGTGGGTTCTCTCCTAGTCAGCCCTTGG 528
```

```
Qy          48 IleGluTyr 50
                  |||
Db         529 TCTGTTTAC 537
```

Search completed: October 26, 2005, 15:25:31
Job time : 106.485 secs

Against DNA

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 03:15:37 ; Search time 170.85 Seconds
(without alignments)
2113.573 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGYLLYXPXMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105431_6380/app_query.fasta_1.917
-DB=N_Geneseq -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954_@CGN_1_1_703_@runat_25102005_105431_6380 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	Query Match %	Length	DB	ID	Description
	1	295	90.5	1514	2	AAV65632	Aav65632 Human hea
	2	295	90.5	5218	13	ACN40746	Acn40746 Tumour-as
	3	295	90.5	5266	10	ADF81754	Adf81754 Leukaemia
	4	295	90.5	11284	2	AAV65633	Aav65633 Plasmid c
	5	232.5	71.3	5230	4	AAK51510	Aak51494 Human pol
	6	220.5	67.6	4850	4	AAK52494	Aak52494 Human pol
	7	204	62.6	5009	2	AAV65634	Aav65634 First Eco
	8	177	54.3	1209	4	ABL13947	Abl13947 Drosophil
c	9	145	44.5	3264	4	ABL13946	Abl13946 Drosophil
c	10	145	44.5	3793	4	ABL13894	Abl13894 Drosophil
	11	76	23.3	760	3	AAZ98210	Aaz98210 Human sig
	12	69.5	21.3	594	8	ACA37736	Aca37736 Prokaryot
c	13	68.5	21.0	42325	10	ADB74382	Adb74382 Mycobacte
	14	67.5	20.7	2720	10	ADB53234	Adb53234 Primary r
	15	66.5	20.4	1660	2	AAZ96329	Aaz96329 S. pneumo
	16	66.5	20.4	2223	10	ABX06561	Abx06561 S. pneumo
	17	66.5	20.4	2292	13	ADR92372	Adr92372 Novel S.
	18	66.5	20.4	2304	3	AAA05535	Aaa05535 Streptoco
	19	66.5	20.4	2304	4	AAS55608	Aas55608 Streptoco
	20	66.5	20.4	2304	4	AAS55863	Aas55863 Streptoco
	21	66.5	20.4	2304	8	ACA49857	Aca49857 Prokaryot
	22	66.5	20.4	2351	2	AAX16136	Aax16136 DNA encod
	23	66.5	20.4	18627	2	AAV52246	Aav52246 Streptoco
	24	66.5	20.4	110000	10	ABS56454_07	Continuation (8 of
c	25	65	19.9	406	10	ADB50839	Adb50839 Primary r
c	26	65	19.9	478	10	ABT40319	Abt40319 Toxicity
c	27	65	19.9	520	10	ADB56829	Adb56829 Toxicity-
	28	65	19.9	97081	12	ADQ97980	Adq97980 Human can
	29	65	19.9	110000	5	AAI61373_0	Aai61373 Soybean 3
	30	65	19.9	110000	5	AAI61373_1	Continuation (2 of
c	31	64.5	19.8	575	12	ADJ75816	Adj75816 Marker ge
	32	64.5	19.8	2155	10	ADI23671	Adi23671 Mouse LPD
	33	64	19.6	1018	8	ACA49002	Aca49002 Prokaryot
	34	64	19.6	1580	8	ADA69987	Ada69987 Rice gene
c	35	64	19.6	4184	4	AAK83286	Aak83286 Human imm
c	36	63.5	19.5	510	12	ACH76865	Ach76865 Human gen
c	37	63.5	19.5	5983	4	AAK84657	Aak84657 Human imm
	38	63	19.3	577	6	ABL93098	Abl93098 Rat metas
	39	63	19.3	577	12	ADN07889	Adn07889 Human mam
	40	63	19.3	8640	4	ABL11790	Abl11790 Drosophil
c	41	62.5	19.2	1314	8	ACA36930	Aca36930 Prokaryot
	42	62.5	19.2	3315	11	ADO58598	Ado58598 Porcine a
c	43	62.5	19.2	73507	11	ACN44612	Acn44612 Mouse gen
	44	62.5	19.2	110000	6	ABA03041_03	Continuation (4 of
c	45	62	19.0	8455	13	ABD33038	Abd33038 Mouse can

ALIGNMENTS

RESULT 1
AAV65632
ID AAV65632 standard; DNA; 1514 BP.
XX
AC AAV65632;
XX
DT 16-DEC-1998 (first entry)
XX
DE Human heart O-fucosyltransferase encoding DNA.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_feature 1. .1100
FT /*tag= b
FT /note= "this actively expressed O-fucosyltransferase
FT sequence is claimed for in claim 9"
FT CDS 1. .1098
FT /*tag= a
FT /product= "human heart O-fucosyltransferase"
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.

XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
DR P-PSDB; AAW80571.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Claim 9; Fig 12A; 90pp; English.
XX
CC This DNA encodes a human heart O-fucosyltransferase that can glycosylate
CC an epidermal growth factor (EGF) domain of a polypeptide with an
CC activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.
CC mutants with increased affinity for the EFG domains, are used in
CC diagnosis and treatment of conditions associated with overexpression of O
CC -fucosyltransferase, to promote survival of sensory (retinal) neurons.
CC Probes based on EGF domain polypeptide are used to detect gene
CC amplification and expression. The expression can also be determined at
CC the protein level using antibodies specific for O-fucosyltransferase
XX
SQ Sequence 1514 BP; 320 A; 435 C; 411 G; 348 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5.87e-33	Length:	1514
Score:	295.00	Matches:	54
Percent Similarity:	91.53%	Conservative:	0
Best Local Similarity:	91.53%	Mismatches:	5
Query Match:	90.49%	Indels:	0
DB:	2	Gaps:	0

US-09-774-954-3 (1-61) x AAV65632 (1-1514)

```
Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db          7 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 66

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db         67 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 126

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 183
```

RESULT 2
ACN40746
ID ACN40746 standard; cDNA; 5218 BP.
XX
AC ACN40746;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) cDNA DNA326813, SEQ ID NO:5725.
XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2004030615-A2.
XX
PD 15-APR-2004.
XX
PF 29-SEP-2003; 2003WO-US028547.
XX
PR 02-OCT-2002; 2002US-0414971P.
XX
PA (GETH) GENENTECH INC.
XX
PI Wu TD, Zhang Z, Zhou Y;
XX
DR WPI; 2004-347921/32.
DR P-PSDB; ABM82223.
XX
PT New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
PS Claim 1; SEQ ID NO 5725; 7273pp; English.
XX
CC The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT nucleic acid of the invention
XX
SQ Sequence 5218 BP; 1255 A; 1274 C; 1334 G; 1355 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.85e-32	Length:	5218
Score:	295.00	Matches:	54
Percent Similarity:	91.53%	Conservative:	0
Best Local Similarity:	91.53%	Mismatches:	5
Query Match:	90.49%	Indels:	0
DB:	13	Gaps:	0

US-09-774-954-3 (1-61) x ACN40746 (1-5218)

```
Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db        125 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 184

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        185 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 244

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 301
```

RESULT 3
ADF81754
ID ADF81754 standard; DNA; 5266 BP.
XX
AC ADF81754;
XX
DT 26-FEB-2004 (first entry)
XX

DE Leukaemia-related DNA sequence #2310.
XX
KW Cytostatic; Gene therapy; leukaemia; ss.
XX
OS Unidentified.
XX
PN WO2003039443-A2.
XX
PD 15-MAY-2003.
XX
PF 04-NOV-2002; 2002WO-EP012303.
XX
PR 05-NOV-2001; 2001EP-00126244.
PR 30-APR-2002; 2002EP-00009758.
XX
PA (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA (UYLU-) UNIV LUDWIG MAXIMILIANS.
PA (HAFE/) HAFERLACH T.
PA (SCHO/) SCHOCH C.
PA (KERN/) KERN W.
XX
PI Haferlach T, Schoch C, Kern W, Kohlmann A, Schnittger S, Dugas M;
PI Eils R, Brors B, Mergenthaler S;
XX
DR WPI; 2003-505037/47.
XX
PT Determining the subtype of leukemia cells and whether a patient sample
PT contains leukemia cells or other cells, useful for treating leukemia,
PT comprises determining the expression profile of a group of markers in a
PT patient sample.
XX
PS Disclosure; SEQ ID NO 2310; 2938pp; English.
XX
CC The present invention relates to a method (M1) for determining the
CC subtype of leukaemia cells and whether a patient sample contains
CC leukaemia cells. The method comprises determining the expression profile
CC of a group of markers in a patient sample. The method is useful for
CC determining the presence of leukaemia cells, its types or subtypes, and
CC for the preparation of a medicament for treating leukaemia.
XX
SQ Sequence 5266 BP; 1280 A; 1278 C; 1331 G; 1346 T; 0 U; 31 Other;

Alignment Scores:

Pred. No.:	2.88e-32	Length:	5266
Score:	295.00	Matches:	54
Percent Similarity:	91.53%	Conservative:	0
Best Local Similarity:	91.53%	Mismatches:	5
Query Match:	90.49%	Indels:	0
DB:	10	Gaps:	0

US-09-774-954-3 (1-61) x ADF81754 (1-5266)

```
Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||  ||||||||||||||||||  |||  ||||||||||||||||
Db        138 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 197

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              |||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        198 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 257

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 314
```

RESULT 4
AAV65633
ID AAV65633 standard; DNA; 11284 BP.
XX
AC AAV65633;
XX
DT 16-DEC-1998 (first entry)
XX
DE Plasmid construct for expression of human O-fucosyltransferase.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS Homo sapiens.
OS Synthetic.
XX

FH	Key	Location/Qualifiers
FT	CDS	4140. .5333
FT		/*tag= a
FT		/note= "insert coding for human O-fucosyltransferase."
FT	misc_feature	4218. .4235
FT		/*tag= b
FT		/note= "polyhistidine tag"

XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
DR P-PSDB; AAW80573.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Example; Page 49-57; 90pp; English.
XX
CC This represents the nucleotide sequence of the plasmid construct used for
CC the expression of human heart O-fucosyltransferase. The human O-
CC fucosyltransferase can glycosylate an epidermal growth factor (EGF)
CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC domains, are used in diagnosis and treatment of conditions associated
CC with overexpression of O-fucosyltransferase, to promote survival of
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC used to detect gene amplification and expression. The expression can also
CC be determined at the protein level using antibodies specific for O-
CC fucosyltransferase
XX
SQ Sequence 11284 BP; 2947 A; 2602 C; 2624 G; 3111 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	7.63e-32	Length:	11284
Score:	295.00	Matches:	54
Percent Similarity:	91.53%	Conservative:	0
Best Local Similarity:	91.53%	Mismatches:	5
Query Match:	90.49%	Indels:	0
DB:	2	Gaps:	0

US-09-774-954-3 (1-61) x AAV65633 (1-11284)

```
Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||  ||||||||||||||||||  |||  ||||||||||||||||
Db       4242 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 4301

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              |||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db       4302 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 4361

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4362 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 4418
```

```
RESULT 5
AAK51510
ID   AAK51510 standard; cDNA; 5230 BP.
XX
AC   AAK51510;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 55.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y;
PI   Ma Y,  Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM78377.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 1; Page 638-642; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 5230 BP; 1285 A; 1255 C; 1330 G; 1360 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                4.11e-23        Length:          5230
Score:                    232.50          Matches:         44
Percent Similarity:       80.00%          Conservative:    4
Best Local Similarity:    73.33%          Mismatches:      7
Query Match:              71.32%          Indels:          5
DB:                       4               Gaps:            1

US-09-774-954-3 (1-61) x AAK51510 (1-5230)

Qy          7 AspLeuAlaGlyTyrLeuLeuTyr***Pro***Met---------------GlyArgPhe 21
                  |||:::|||   :::      :::   |||   :::                 |||||||||
Db        160 GACATGGCCACGTGGCCTACACATGCTCCAAGCGTGGAGGAAGGAGGCCAGGGGCGCTTT 219

Qy         22 GlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThr 41
               ||||||||||||||||||||||||||||||||||||||||||||||||||      ||||||
Db        220 GGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACC 279

Qy         42 LeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 TTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 339

RESULT 6
AAK52494
ID   AAK52494 standard; cDNA; 4850 BP.
XX
AC   AAK52494;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 2023.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y;
PI   Ma Y,  Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   P-PSDB; AAM79361.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 1; Page 4424-4425; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 4850 BP; 1142 A; 1218 C; 1244 G; 1246 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:              2.14e-21        Length:         4850
Score:                  220.50          Matches:        42
Percent Similarity:     78.33%          Conservative:   5
Best Local Similarity:  70.00%          Mismatches:     8
Query Match:            67.64%          Indels:         5
DB:                     4               Gaps:           1

US-09-774-954-3 (1-61) x AAK52494 (1-4850)

Qy          7 AspLeuAlaGlyTyrLeuLeuTyr***Pro***Met----------GlyArgPhe 21
              |||:::|||   :::     :::   |||   :::              ||||||||
Db        160 GACATGGCCACGTGGCCTACACATGCTCCAAGCGTGGAGGAAGGAGGCCAGGGGCGCTTT 219

Qy         22 GlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThr 41
              |||||||||||||||||||||||||||||||||||||||||||||||||     |||:::
Db        220 GGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTGAACCGTTCC 279

Qy         42 LeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db        280 TTGGCTGTCCCTTCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 339

RESULT 7
AAV65634
ID   AAV65634 standard; cDNA; 5009 BP.
XX
AC   AAV65634;
XX
DT   16-DEC-1998  (first entry)
XX
DE   First EcoR1 nucleotide fragment of human KIAA0180.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS   Homo sapiens.
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
DR   P-PSDB; AAW80577.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example 1; Fig 11; 90pp; English.
XX
CC   This represents a first EcoR1 nucleotide fragment of human KIAA0180. This
CC   5009 basepairs partial cDNA encodes for a protein of unknown function
CC   from myeblast cel line KG-1. The invention provides a human heart O-
CC   fucosyltransferase that can glycosylate an epidermal growth factor (EGF)
CC   domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC   O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 5009 BP; 1235 A; 1195 C; 1268 G; 1311 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              5.84e-19        Length:         5009
Score:                  204.00          Matches:        37
Percent Similarity:     94.87%          Conservative:   0
Best Local Similarity:  94.87%          Mismatches:     2
Query Match:            62.58%          Indels:         0
DB:                     2               Gaps:           0

US-09-774-954-3 (1-61) x AAV65634 (1-5009)

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              |||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db          2 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 61

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 118

RESULT 8
ABL13947
ID   ABL13947 standard; cDNA; 1209 BP.
XX
AC   ABL13947;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36323.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69844.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 36323; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1209 BP; 302 A; 313 C; 307 G; 287 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              8.6e-16         Length:         1209
Score:                  177.00          Matches:        33
Percent Similarity:     80.00%          Conservative:   3
```

```
Best Local Similarity:  73.33%          Mismatches:     9
Query Match:            54.29%          Indels:         0
DB:                     4               Gaps:           0

US-09-774-954-3 (1-61) x ABL13947 (1-1209)

Qy           7 AspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyAsnGlnAlaAsp 26
               |||       |||||||||    |||   |||   ||||||||||||||||||||||||
Db          82 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCGAC 141

Qy          27 HisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProPro 46
               ||||||||||||||||||||||||||||||          |||||||||   :::|||||| 
Db         142 CACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCG 201

Qy          47 TrpIleGluTyrGln 51
               |||:::||||||:::
Db         202 TGGGTGGAGTATCGT 216

RESULT 9
ABL13946/c
ID   ABL13946 standard; cDNA; 3264 BP.
XX
AC   ABL13946;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36320.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69843.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 36320; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3264 BP; 904 A; 742 C; 758 G; 860 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.49e-10        Length:         3264
Score:                  145.00          Matches:        26
Percent Similarity:     87.88%          Conservative:   3
Best Local Similarity:  78.79%          Mismatches:     4
Query Match:            44.48%          Indels:         0
DB:                     4               Gaps:           0

US-09-774-954-3 (1-61) x ABL13946 (1-3264)

Qy          19 GlyArgPheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu*** 38
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        2092 GGACGCTTTGGCAACCAGGCCGACCACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTT 2033

Qy          39 ValArgThrLeuAlaValProProTrpIleGluTyrGln 51
               |||||||||   :::|||||||||:::||||||:::
Db        2032 AATCGCACCCTGATCCTGCCGCCGTGGGTGGAGTATCGT 1994

RESULT 10
ABL13894/c
ID   ABL13894 standard; cDNA; 3793 BP.
XX
AC   ABL13894;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36164.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69791.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 36164; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3793 BP; 1020 A; 853 C; 950 G; 970 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.81e-10        Length:         3793
Score:                  145.00          Matches:        26
Percent Similarity:     87.88%          Conservative:   3
Best Local Similarity:  78.79%          Mismatches:     4
Query Match:            44.48%          Indels:         0
DB:                     4               Gaps:           0

US-09-774-954-3 (1-61) x ABL13894 (1-3793)
```

```
Qy         19 GlyArgPheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu*** 38
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        584 GGACGCTTTGGCAACCAGGCCGACCACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTT 525

Qy         39 ValArgThrLeuAlaValProProTrpIleGluTyrGln 51
                 |||||||||   :::|||||||||:::||||||:::
Db        524 AATCGCACCCTGATCCTGCCGCCGTGGGTGGAGTATCGT 486

RESULT 11
AAZ98210
ID   AAZ98210 standard; cDNA; 760 BP.
XX
AC   AAZ98210;
XX
DT   11-MAY-2000  (first entry)
XX
DE   Human signal peptide containing protein HSPP-102 cDNA SEQ ID NO:236.
XX
KW   Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
KW   inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW   antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;
KW   antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW   reproductive disorder; developmental disorder; arteriosclerosis;
KW   cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW   asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW   Parkinson's disease; Huntington's diseases; ovulatory defect;
KW   muscular dystrophy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200000610-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US014484.
XX
PR   26-JUN-1998;   98US-0090762P.
PR   31-JUL-1998;   98US-0094983P.
PR   01-OCT-1998;   98US-0102686P.
PR   11-DEC-1998;   98US-0112129P.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Lal P,  Tang YT,  Gorgone GA,  Corley NC,  Guegler KJ,  Baughn MR;
PI   Akerblom IE,  Au-Young J,  Yue H,  Patterson C,  Reddy R,  Hillman JL;
PI   Bandman O;
XX
DR   WPI; 2000-160673/14.
DR   P-PSDB; AAY87325.
XX
PT   New human signal peptide-containing proteins useful in treatment,
PT   prevention and diagnosis of e.g. cancer, inflammation and cardiovascular
PT   disease.
XX
PS   Claim 9; Page 307; 327pp; English.
XX
CC   AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the
CC   human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
CC   anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,
CC   neuroprotective, cardiovascular and antiasthmatic activities, and can be
CC   used in gene therapy. HSPPs can be used to treat or prevent disorders
CC   associated with decreased activity or function of HSPP. Antagonists of
CC   HSPP are used to treat or prevent disorders associated with increased
CC   activity or function of HSPP. Such diseases include cell proliferation
CC   (including cancer), inflammation, cardiovascular, neurological,
CC   reproductive or developmental disorders, (e.g. arteriosclerosis,
CC   cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
CC   asthma, Crohn's disease, microbial or other infections, congestive or
CC   ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
CC   diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
CC   nucleic acids can be used for the recombinant production of HSPP, for
CC   detecting HSPP in standard hybridisation and amplification assays (for
CC   diagnosis and monitoring), in gene therapy, as antisense, triplex-forming
CC   or ribozyme therapeutics, for detecting related sequences or genetic
CC   variations, and for chromosomal mapping. HSPP are also used to raise
CC   specific antibodies (Ab) and to screen for agonists and antagonists
CC   (potential therapeutic agents). Ab are used to diagnose, or monitor, HSPP
CC   -related diseases (in usual immunoassays), as therapeutic antagonists, in
CC   competitive drug screens, and for purification of HSPP from natural
CC   sources
XX
SQ   Sequence 760 BP; 171 A; 206 C; 220 G; 163 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                0.3          Length:          760
Score:                    76.00        Matches:         14
Percent Similarity:       83.33%       Conservative:    1
Best Local Similarity:    77.78%       Mismatches:      3
Query Match:              23.31%       Indels:          0
DB:                       3            Gaps:            0

US-09-774-954-3 (1-61) x AAZ98210 (1-760)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              ||||||||||||||||   ||||||||||||||||||   |||   ||||||:::
Db        117 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGTAAG 170

RESULT 12
ACA37736
ID   ACA37736 standard; DNA; 594 BP.
XX
AC   ACA37736;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Prokaryotic essential gene #19393.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
OS   Mycobacterium avium.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   P-PSDB; ABU33866.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 14; SEQ ID NO 25606; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
```

```
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC   prokaryotic essential genes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 594 BP; 67 A; 204 C; 238 G; 85 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                1.96          Length:        594
Score:                    69.50         Matches:       16
Percent Similarity:       56.25%        Conservative:  2
Best Local Similarity:    50.00%        Mismatches:    11
Query Match:              21.32%        Indels:        3
DB:                       8             Gaps:          1

US-09-774-954-3 (1-61) x ACA37736 (1-594)

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              ||||||   |||   |||   |||   |||   :::||||||||||   |||   |||
Db        346 TTTGGACGCCAGGTCGACTCGTTCGAGGGTGACATCGCGTTCGCGGGACTGGACGTGCCG 405

Qy         41 ThrLeuAlaVal---------ProProTrpIleGlu 49
                  ||||||          ||||||:::|||
Db        406 GTGCGCGCGGTGTTCATCCGCGCGCCCTGGGTGGAG 441

RESULT 13
ADB74382/c
ID   ADB74382 standard; DNA; 42325 BP.
XX
AC   ADB74382;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Mycobacterium leprae DNA #16.
XX
KW   Non-naturally occurring peptide; anion pump protein; tuberculosis;
KW   hypersensitivity reaction; tuberculostatic; gene; ds.
XX
OS   Mycobacterium leprae.
XX
PN   US6583266-B1.
XX
PD   24-JUN-2003.
XX
PF   16-SEP-1994;   94US-00311731.
XX
PR   19-AUG-1993;   93US-00109181.
PR   22-OCT-1993;   93US-00142558.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Smith DR,  Mao J;
XX
DR   WPI; 2003-656441/62.
XX
PT   New Mycobacterium tuberculosis anion pump peptide useful for as
PT   tuberculosis vaccine and diagnosis of tuberculosis infection.
XX
PS   Disclosure; SEQ ID NO 131; 26pp; English.
XX
CC   The invention relates to a non-naturally occurring peptide of
CC   Mycobacterium tuberculosis comprising an amino acid sequence
CC   corresponding to an anion pump protein. The invention also relates to a
CC   non-naturally occurring nucleic acid corresponding to a DNA sequence of
CC   Mycobacterium tuberculosis or Mycobacterium leprae. The new peptide is
CC   useful as a vaccine against Mycobacterium tuberculosis or Mycobacterium
CC   leprae or for screening for new tuberculosis drugs. Purified proteins
CC   derived from the sequences of the invention may elicit a specific immune
CC   response. The peptide may also be used to detect hypersensitivity
CC   reactions of individuals exposed to Mycobacterium tuberculosis or
CC   Mycobacterium leprae. The proteins and peptides may be affixed to solid
CC   supports to detect antibodies typical of hypersensitivity reactions, from
CC   a patient's sera. This sequence represents Mycobacterium leprae DNA of
CC   the invention. Note: The sequence data for this patent did not form part
CC   of the printed specification but was obtained in electronic format
CC   directly from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 42325 BP; 9673 A; 13128 C; 11330 G; 8194 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                639           Length:        42325
Score:                    68.50         Matches:       23
Percent Similarity:       42.65%        Conservative:  6
Best Local Similarity:    33.82%        Mismatches:    14
Query Match:              21.01%        Indels:        25
DB:                       10            Gaps:          4

US-09-774-954-3 (1-61) x ADB74382 (1-42325)

Qy          5 SerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyAsnGln 24
              ||||||||||||      :::   |||:::   |||   :::|||   :::|||   |||
Db      30298 TCCTGGGATCTGCGTAAATTC---CTCTGGGCGCCGAAGCTGGGCTGCTACGGCATCCAA 30242

Qy         25 AlaAspHisPheLeu----------------------------------------GlySer 31
                  |||   |||                                            ||||||
Db      30241 CGCATTCACCTGCTGCGCAACGTGATGATATTGGCCGGCGCCGGAGTGGGAGGCGGTTCG 30182

Qy         32 LeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProProTrpIleGluTyrGln 51
              |||   :::|||                |||||   ||||||||
Db      30181 TTGAACTACGCG------------AATACGTTGTACGTTCCGCCG--------------- 30149

Qy         52 HisHisLysProProPheThrAsn 59
                 :::|||   |||   |||
Db      30148 ------GAGCCGTTCTTTGCCAAC 30131

RESULT 14
ADB53234
ID   ADB53234 standard; DNA; 2720 BP.
XX
AC   ADB53234;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:3776.
XX
KW   toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
KW   toxicity marker; toxicity progression; drug screening;
KW   primary rat hepatocyte toxicity modelling; gene; ds.
XX
OS   Rattus norvegicus.
XX
PN   WO2003065993-A2.
XX
PD   14-AUG-2003.
XX
PF   04-FEB-2003; 2003WO-US003482.
XX
```

PR 04-FEB-2002; 2002US-0353171P.
PR 13-MAR-2002; 2002US-0363534P.
PR 08-APR-2002; 2002US-0370248P.
PR 10-APR-2002; 2002US-0371134P.
PR 10-APR-2002; 2002US-0371135P.
PR 10-APR-2002; 2002US-0371150P.
PR 11-APR-2002; 2002US-0371413P.
PR 19-APR-2002; 2002US-0373601P.
PR 19-APR-2002; 2002US-0373602P.
PR 22-APR-2002; 2002US-0374139P.
PR 08-MAY-2002; 2002US-0378370P.
PR 09-MAY-2002; 2002US-0378652P.
PR 09-MAY-2002; 2002US-0378653P.
PR 09-MAY-2002; 2002US-0378665P.
PR 09-JUL-2002; 2002US-0394230P.
PR 09-JUL-2002; 2002US-0394253P.
PR 04-SEP-2002; 2002US-0407688P.
PR 28-JAN-2003; 2003US-0442900P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;
PI Elashoff M;
XX
DR WPI; 2003-731472/69.
XX
PT Determining if a compound induces a toxic effect on a tissue or cell, for
PT identifying hepatotoxic compounds, comprises comparing a gene expression
PT profile of a tissue or cell sample to a database of Tox mean and non-Tox
PT mean values.
XX
PS Claim 44; SEQ ID NO 3776; 874pp; English.
XX
CC The present invention describes a method for determining whether a
CC compound induces a toxic effect on a tissue or cell. The method comprises
CC preparing a gene expression profile of a tissue or cell sample exposed to
CC the compound, and comparing the gene expression profile to a database
CC comprising data or information on the Tox mean and non-Tox mean value.
CC The method is useful for predicting or identifying at least one toxic
CC effect, particularly hepatotoxicity, of a test or unknown compound. The
CC genes listed in the specification are useful as diagnostic or toxicity
CC markers for the prediction or identification of the physiological state
CC of tissue or cell sample that has been exposed to a compound, or to
CC identify or predict the toxic effects of a compound or an agent. These
CC may also be used as markers for monitoring toxicity progression or for
CC drug screening. The present sequence represents a primary rat hepatocyte
CC toxicity modelling related gene sequence from the present invention.
XX
SQ Sequence 2720 BP; 624 A; 719 C; 725 G; 652 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	26.9	Length:	2720
Score:	67.50	Matches:	26
Percent Similarity:	34.74%	Conservative:	7
Best Local Similarity:	27.37%	Mismatches:	19
Query Match:	20.71%	Indels:	43
DB:	10	Gaps:	5

US-09-774-954-3 (1-61) x ADB53234 (1-2720)

```
Qy          2 LeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPhe 21
              |||:::|||:::|||   :::              |||      |||||||||
Db        713 CTCTCTGGGACTTGGACAATC-----------------ACCCCCGGA---GGCAGGTTT 751

Qy         22 GlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg--- 40
              |||||||||         :::         ||||||   |||:::|||      |||
Db        752 GGTAACCAGATGGGACAGTATGCCACATTGCTGGCCCTAGCCCAGCTCAATGGCCGCCGG 811

Qy         41 --------------------------ThrLeuAla---------------------- 43
                                        |||||||||
Db        812 GCTTTCATCCAGCCTGAGATGCATACCACCCTGGCCCCTGTGTTCCGAATCTCCCTGCCA 871

Qy         44 ----------------------ValProProTrp---------------------- 47
```

```
                                          :::   ||||||
Db        872 GTGCTGGACCCTGAGGTGGACAGTCTCACGCCTTGGCAGCACTTAGTCCTACATGACTGG 931

Qy         48 ------IleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
                    ||||||   |||   :::   ||||||      |||
Db        932 ATGTCAGAGGAGTACTCCCACCTGGAGGACCCGTTTCTCAAGCTG 976
```

RESULT 15
AAZ96329
ID AAZ96329 standard; DNA; 1660 BP.
XX
AC AAZ96329;
XX
DT 10-APR-2000 (first entry)
XX
DE S. pneumoniae derived DNA from ORF #157.
XX
KW Treatment; prevention; disease; diagnosis; gene therapy; screening;
KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.
XX
OS Streptococcus pneumoniae.
XX
PN WO9806734-A1.
XX
PD 19-FEB-1998.
XX
PF 15-AUG-1997; 97WO-US014436.
XX
PR 16-AUG-1996; 96US-0024022P.
XX
PA (SMIK) SMITHKLINE BEECHAM CORP.
XX
PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Stodola RK;
XX
DR WPI; 1998-159452/14.
DR P-PSDB; AAY85984.
XX
PT Streptococcus pneumoniae proteins and related DNA - useful for screening
PT compounds for antibacterial activity.
XX
PS Claim 4; Page 191-192; 640pp; English.
XX
CC This invention describes novel isolated Streptococcus pneumoniae
CC polynucleotides (see AAZ96173-Z96494) and their encoded proteins (see
CC AAY85792-Y86182). The DNA, vectors and host cells described in the method
CC of the invention are useful for the recombinant expression of the
CC polypeptides. The polypeptides are useful for treatment or prevention of
CC disease, or diagnosis of disease related to expression or activity of
CC such a polypeptide. They can also be used to screen for compounds which
CC interact with and inhibit or activate such a polypeptide. The
CC polypeptides (or DNA encoding them, via gene therapy) are also useful for
CC inducing an immunological response in a mammal. The antagonists are
CC useful to inhibit such bacterial polypeptides. The polypeptides are
CC particularly useful to identify antimicrobial compounds and antibiotics.
CC They are also useful to determine their role in pathogenesis of
CC infection, dysfunction and disease
XX
SQ Sequence 1660 BP; 467 A; 332 C; 379 G; 479 T; 0 U; 3 Other;

Alignment Scores:

Pred. No.:	20.1	Length:	1660
Score:	66.50	Matches:	16
Percent Similarity:	48.84%	Conservative:	5
Best Local Similarity:	37.21%	Mismatches:	19
Query Match:	20.40%	Indels:	3
DB:	2	Gaps:	1

US-09-774-954-3 (1-61) x AAZ96329 (1-1660)

```
Qy          8 LeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGlyAsnGlnAlaAspHis 27
              |||   |||      ||||||   |||      |||   |||        :::
Db       1077 TTGATCGGGGTCGCTCTTTATATTCCAACAGCCTTTCTCTTTTCAAATATCGGAACTTAC 1136
```

```
Qy          28 PheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeuAlaValProProTrp 47
               |||:::||||||:::           :::   |||   :::|||   |||   ||||||
Db        1137 TTTATTGGTTCTATCTTG---------ATTTTAGTGGGTTCTCTCCTAGTCAGCCCTTGG 1187

Qy          48 IleGluTyr 50
               |||
Db        1188 TCTGTTTAC 1196


Search completed: October 26, 2005, 09:51:41
Job time : 185.85 secs
```

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 05:08:04 ; Search time 1382.21 Seconds
(without alignments)
2138.440 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGYLLYXPXMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105432_6389/app_query.fasta_1.917
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954_@CGN_1_1_4968_@runat_25102005_105432_6389 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	305	93.6	1302	10	AF375885	AF375885 Mus muscu
2	305	93.6	1529	10	BC046295	BC046295 Mus muscu
3	302	92.6	1188	10	AJ781499	AJ781499 Rattus no
4	295	90.5	1100	6	AR163457	AR163457 Sequence
5	295	90.5	1100	6	BD103196	BD103196 O-fucosyl
6	295	90.5	1167	6	CQ727777	CQ727777 Sequence
7	295	90.5	1167	9	AJ781500	AJ781500 Pan trogl
8	295	90.5	1300	6	AR163452	AR163452 Sequence
9	295	90.5	1300	6	BD103189	BD103189 O-fucosyl
10	295	90.5	1514	6	AR163451	AR163451 Sequence
11	295	90.5	1514	6	BD103188	BD103188 O-fucosyl
12	295	90.5	1524	9	BC000582	BC000582 Homo sapi
13	295	90.5	4560	9	HSM803330	AL832023 Homo sapi
14	295	90.5	5189	9	D80002	D80002 Homo sapien
15	295	90.5	5249	9	AF375884	AF375884 Homo sapi
16	295	90.5	5266	6	AX780153	AX780153 Sequence
17	295	90.5	11284	6	BD103190	BD103190 O-fucosyl
18	288	88.3	1056	4	AY344581	AY344581 Bos tauru
19	288	88.3	1176	4	AY344580	AY344580 Bos tauru
20	286	87.7	1134	4	SSC567917	AJ567917 Sus scrof
21	273	83.7	1176	4	AJ781503	AJ781503 Bos tauru
22	272	83.4	920	5	AJ719585	AJ719585 Gallus ga
23	272	83.4	1732	5	GGA535754	AJ535754 Gallus ga
24	272	83.4	2546	5	AJ720352	AJ720352 Gallus ga
25	250	76.7	1502	5	BC082519	BC082519 Xenopus t
26	250	76.7	1538	5	AJ781498	AJ781498 Silurana
27	249	76.4	1549	5	DRE579536	AJ579536 Danio rer
28	247	75.8	1155	5	AJ606070	AJ606070 Fugu rubr
29	247	75.8	1164	5	AJ781504	AJ781504 Tetraodon
30	234	71.8	1505	5	XLA514425	AJ514425 Xenopus l
31	220	67.5	145414	9	HSA392M18	AL121897 Human DNA
32	220	67.5	200329	10	AC078911	AC078911 Mus muscu
33	220	67.5	208614	10	AL807380	AL807380 Mouse DNA
c 34	220	67.5	240336	2	AC094844	AC094844 Rattus no
c 35	220	67.5	240510	2	AC134371	AC134371 Rattus no
36	204	62.6	5009	6	AR163453	AR163453 Sequence
37	204	62.6	5009	6	BD103191	BD103191 O-fucosyl
38	177	54.3	1209	3	AJ831490	AJ831490 Drosophil
39	177	54.3	1209	6	CQ590403	CQ590403 Sequence
40	177	54.3	1295	3	AY118651	AY118651 Drosophil
41	177	54.3	1320	6	AR509393	AR509393 Sequence
42	177	54.3	1579	3	AB093572	AB093572 Drosophil
43	170	52.1	1146	3	AJ781502	AJ781502 Ciona sav
44	164	50.3	1209	3	AJ781501	AJ781501 Drosophil
45	162	49.7	1841	3	AK112708	AK112708 Ciona int

ALIGNMENTS

RESULT 1
AF375885
LOCUS AF375885 1302 bp mRNA linear ROD 23-OCT-2001
DEFINITION Mus musculus protein o-fucosyltransferase (Pofut1) mRNA, complete cds.
ACCESSION AF375885
VERSION AF375885.1 GI:15825115
KEYWORDS .
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1302)
AUTHORS Wang,Y., Shao,L., Shi,S., Harris,R.J., Spellman,M.W., Stanley,P. and Haltiwanger,R.S.
TITLE Modification of epidermal growth factor-like repeats with O-fucose. Molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase
JOURNAL J. Biol. Chem. 276 (43), 40338-40345 (2001)
MEDLINE 21523965
PUBMED 11524432
REFERENCE 2 (bases 1 to 1302)
AUTHORS Shi,S., Stanley,P., Wang,Y., Shao,L., Harris,R.J., Spellman,M.W. and Haltiwanger,R.S.
TITLE Direct Submission
JOURNAL Submitted (02-MAY-2001) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave., New York, NY 10461, USA
FEATURES Location/Qualifiers

```
     source          1. .1302
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="129/SvJ"
                     /db_xref="taxon:10090"
     gene            1. .1302
                     /gene="Pofut1"
     CDS             25. .1206
                     /gene="Pofut1"
                     /note="O-FucT-1"
                     /codon_start=1
                     /product="protein o-fucosyltransferase"
                     /protein_id="AAL09577.1"
                     /db_xref="GI:15825116"
                     /translation="MGAAAWAPPHLLLRASFLLLLLLLPLRGRSAGSWDLAGYLLYCP
                     CMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQ
                     AYHRVVSLEDFMENLAPSHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQF
                     HVSFNKSELFTGISFSASYKEQWTQRFPAKEHPVLALPGAPAQFPVLEEHRELQKYMV
                     WSDEMVRTGEALISAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYS
                     RSTATPLTMTMCLPDLKEIQRAVTLWVRALNARSVYIATDSESYVSEIQQLFKDKVRV
                     VSLKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLHGRQSSFFGMDRPSQLRDE
                     F"
ORIGIN

Alignment Scores:
Pred. No.:              1.2e-33         Length:         1302
Score:                  305.00          Matches:        56
Percent Similarity:     91.80%          Conservative:   0
Best Local Similarity:  91.80%          Mismatches:     5
Query Match:            93.56%          Indels:         0
DB:                     10              Gaps:           0

US-09-774-954-3 (1-61) x AF375885 (1-1302)

Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   |||||||||||||||||||||||||||||||||||   |||   |||||||||
Db        109 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 168

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        169 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 228

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 288

Qy         61 His 61
              |||
Db        289 CAT 291

RESULT 2
BC046295
LOCUS       BC046295               1529 bp    mRNA    linear   ROD 30-JUN-2004
DEFINITION  Mus musculus protein O-fucosyltransferase 1, mRNA (cDNA clone
            MGC:54837 IMAGE:6485896), complete cds.
ACCESSION   BC046295
VERSION     BC046295.1  GI:28279489
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1529)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1529)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 100 Row: m Column: 11
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 17978277.
FEATURES            Location/Qualifiers
     source         1. .1529
                    /organism="Mus musculus"
                    /mol_type="mRNA"
                    /strain="FVB/N-3"
                    /db_xref="taxon:10090"
                    /clone="MGC:54837 IMAGE:6485896"
                    /tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month
                    old mouse. Taken by biopsy."
                    /clone_lib="NCI_CGAP_Mam2"
                    /lab_host="DH10B"
                    /note="Vector: pCMV-SPORT6"
     gene           1. .1529
                    /gene="Pofut1"
                    /note="synonym: O-FucT-1"
                    /db_xref="LocusID:140484"
                    /db_xref="MGI:2153207"
     CDS            56. .1237
                    /gene="Pofut1"
                    /codon_start=1
                    /product="protein O-fucosyltransferase 1"
                    /protein_id="AAH46295.1"
                    /db_xref="GI:28279490"
                    /db_xref="LocusID:140484"
                    /db_xref="MGI:2153207"
                    /translation="MGAAAWAPPHLLLRASFLLLLLLLPLRGRSAGSWDLAGYLLYCP
                    CMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQ
                    AYHRVVSLEDFMENLAPSHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQF
                    HVSFNKSELFTGISFSASYKEQWTQRFPAKEHPVLALPGAPAQFPVLEEHRELQKYMV
```

```
                     WSDEMVRTGEALISAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYS
                     RSTATPLTMTMCLPDLKEIQRAVTLWVRALNARSVYIATDSESYVSEIQQLFKDKVRV
                     VSLKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLHGRQSSFFGMDRPSQLRDE
                     F"
ORIGIN

Alignment Scores:
Pred. No.:              1.43e-33        Length:         1529
Score:                  305.00          Matches:        56
Percent Similarity:     91.80%          Conservative:   0
Best Local Similarity:  91.80%          Mismatches:     5
Query Match:            93.56%          Indels:         0
DB:                     10              Gaps:           0

US-09-774-954-3 (1-61) x BC046295 (1-1529)

Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   |||||||||||||||||||||||||||||||||||||    |||   |||||||||
Db        140 CGGTCAGCGGGCTCCTGGGACCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 199

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        200 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCCCTGGCATTTGCGAAGCTGCTGAACCGC 259

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 ACCTTGGCTGTACCTCCATGGATTGAATACCAACATCACAAGCCTCCTTTCACCAACCTC 319

Qy         61 His 61
              |||
Db        320 CAT 322

RESULT 3
AJ781499
LOCUS       AJ781499               1188 bp    mRNA    linear   ROD 07-JUL-2004
DEFINITION  Rattus norvegicus mRNA for protein-O-fucosyltransferase 1 (fut12
            gene).
ACCESSION   AJ781499
VERSION     AJ781499.1  GI:50057067
KEYWORDS    fut12 gene; pofut1 gene; protein-O-fucosyltransferase 1.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1
  AUTHORS   Martinez-Duncker,I., Mollicone,R., Candelier,J.J., Breton,C. and
            Oriol,R.
  TITLE     A new superfamily of protein-O-fucosyltransferases,
            alpha2-fucosyltransferases, and alpha6-fucosyltransferases:
            phylogeny and identification of conserved peptide motifs
  JOURNAL   Glycobiology 13 (12), 1C-5C (2003)
   PUBMED   12966037
REFERENCE   2
  AUTHORS   Martinez-Duncker,I., Oriol,R. and Mollicone,R.
  TITLE     Phylogeny of fucosyltransferases
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1188)
  AUTHORS   Oriol,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Oriol R., U504, Inserm, 16 Av. Paul
            Vaillant-Couturier, 94807, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1188
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Wistar"
                     /db_xref="taxon:10116"
     gene            1. .1188
                     /gene="fut12"
     CDS             1. .1188
                     /gene="fut12"
                     /function="adds fucose to Ser or Thr"
                     /note="pofut1"
                     /codon_start=1
                     /product="protein-O-fucosyltransferase 1"
                     /protein_id="CAH03711.1"
                     /db_xref="GI:50057068"
                     /translation="MGAAAWAPSHLLLRASFLLLLLLLPLPLRGRAGGSWDLAGYLLY
                     CPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEP
                     LQAYHRVISLEDFMEKLAPFHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWD
                     QFHVSFNKSELFTGISFSASYKEQWIQRFPPKEHPVLALPGAPAQFPVLEEHRELQKY
                     MVWSDEMVRTGEAQISTHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVG
                     YSRSTATPLTTTMCLPDLKEIQRAVKLWVRALDARSVYIATDSESYVSEIQQLFKEKV
                     KVVSLKPEVAQIDLYILGQADHFIGNCVSSFTAFVKRERDLHGRPSSFFGMDRPSQLR
                     DEF"
ORIGIN

Alignment Scores:
Pred. No.:              2.91e-33        Length:         1188
Score:                  302.00          Matches:        55
Percent Similarity:     90.16%          Conservative:   0
Best Local Similarity:  90.16%          Mismatches:     6
Query Match:            92.64%          Indels:         0
DB:                     10              Gaps:           0

US-09-774-954-3 (1-61) x AJ781499 (1-1188)

Qy          1 ArgLeuAlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArg 20
              |||   |||||||||||||||||||||||||||||||||||||    |||   |||||||||
Db         91 CGGGCGGGGGGCTCCTGGGATCTGGCCGGTTACCTGCTCTACTGTCCCTGCATGGGGCGC 150

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||      |||
Db        151 TTTGGGAACCAGGCTGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTGAACCGC 210

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 ACCTTGGCTGTACCTCCATGGATTGAGTACCAACATCACAAGCCTCCCTTCACCAACCTC 270

Qy         61 His 61
              |||
Db        271 CAT 273

RESULT 4
AR163457
LOCUS       AR163457               1100 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 16 from patent US 6270987.
ACCESSION   AR163457
VERSION     AR163457.1  GI:16234062
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: US 6270987-A 16 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              2.69e-32        Length:         1100
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     6               Gaps:           0

US-09-774-954-3 (1-61) x AR163457 (1-1100)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||
```

```
Db          7 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 66

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db         67 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 126

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 183

RESULT 5
BD103196
LOCUS       BD103196                1100 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103196
VERSION     BD103196.1  GI:22648770
KEYWORDS    JP 2001527389-A/9.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 9 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/9
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .1100
            FT                   /organism='Unidentified'.
FEATURES              Location/Qualifiers
     source           1. .1100
                      /organism="unidentified"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:32644"
ORIGIN

Alignment Scores:
Pred. No.:                2.69e-32        Length:          1100
Score:                    295.00          Matches:         54
Percent Similarity:       91.53%          Conservative:    0
Best Local Similarity:    91.53%          Mismatches:      5
Query Match:              90.49%          Indels:          0
DB:                       6               Gaps:            0

US-09-774-954-3 (1-61) x BD103196 (1-1100)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||  ||||||||||||||||||  |||  ||||||||||||||||
Db          7 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 66

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db         67 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 126

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 183

RESULT 6
CQ727777
LOCUS       CQ727777                1167 bp    DNA     linear   PAT 03-FEB-2004
DEFINITION  Sequence 13711 from Patent WO02068579.
ACCESSION   CQ727777
VERSION     CQ727777.1  GI:42294771
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 13711 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES              Location/Qualifiers
     source           1. .1167
                      /organism="Homo sapiens"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                2.87e-32        Length:          1167
Score:                    295.00          Matches:         54
Percent Similarity:       91.53%          Conservative:    0
Best Local Similarity:    91.53%          Mismatches:      5
Query Match:              90.49%          Indels:          0
DB:                       6               Gaps:            0

US-09-774-954-3 (1-61) x CQ727777 (1-1167)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||  ||||||||||||||||||  |||  ||||||||||||||||
Db         76 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 135

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        136 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 195

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 252

RESULT 7
AJ781500
LOCUS       AJ781500                1167 bp    mRNA    linear   PRI 07-JUL-2004
DEFINITION  Pan troglodytes mRNA for protein-O-fucosyltransferase 1 (fut12
            gene).
ACCESSION   AJ781500
VERSION     AJ781500.1  GI:50057069
KEYWORDS    fut12 gene; pofut1 gene; protein-O-fucosyltransferase 1.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Martinez-Duncker,I., Mollicone,R., Candelier,J.J., Breton,C. and
            Oriol,R.
  TITLE     A new superfamily of protein-O-fucosyltransferases,
            alpha2-fucosyltransferases, and alpha6-fucosyltransferases:
            phylogeny and identification of conserved peptide motifs
  JOURNAL   Glycobiology 13 (12), 1C-5C (2003)
   PUBMED   12966037
REFERENCE   2
  AUTHORS   Martinez-Duncker,I., Oriol,R. and Mollicone,R.
  TITLE     Phylogeny of fucosyltransferases
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1167)
  AUTHORS   Oriol,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-JUL-2004) Oriol R., U504, Inserm, 16 Av. Paul
            Vaillant-Couturier, 94807, FRANCE
FEATURES              Location/Qualifiers
     source           1. .1167
                      /organism="Pan troglodytes"
```

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9598"
     gene            1. .1167
                     /gene="fut12"
     CDS             1. .1167
                     /gene="fut12"
                     /function="adds fucose to Ser or Thr"
                     /note="pofut1"
                     /codon_start=1
                     /product="protein-O-fucosyltransferase 1"
                     /protein_id="CAH03712.1"
                     /db_xref="GI:50057070"
                     /translation="MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRF
                     GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRV
                     ISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFN
                     KSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEM
                     VKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA
                     PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKP
                     EVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF"
ORIGIN

Alignment Scores:
Pred. No.:                2.87e-32        Length:         1167
Score:                    295.00          Matches:        54
Percent Similarity:       91.53%          Conservative:   0
Best Local Similarity:    91.53%          Mismatches:     5
Query Match:              90.49%          Indels:         0
DB:                       9               Gaps:           0

US-09-774-954-3 (1-61) x AJ781500 (1-1167)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db         76 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 135

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        136 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 195

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 252

RESULT 8
AR163452
LOCUS       AR163452                1300 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 6 from patent US 6270987.
ACCESSION   AR163452
VERSION     AR163452.1  GI:16234057
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1300)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: US 6270987-A 6 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1300
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:                3.24e-32        Length:         1300
Score:                    295.00          Matches:        54
Percent Similarity:       91.53%          Conservative:   0
Best Local Similarity:    91.53%          Mismatches:     5
Query Match:              90.49%          Indels:         0
DB:                       6               Gaps:           0

US-09-774-954-3 (1-61) x AR163452 (1-1300)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db        142 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 201

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        202 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 261

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 318

RESULT 9
BD103189
LOCUS       BD103189                1300 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103189
VERSION     BD103189.1  GI:22648763
KEYWORDS    JP 2001527389-A/2.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1300)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 2 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/2
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .1300
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1300
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

Alignment Scores:
Pred. No.:                3.24e-32        Length:         1300
Score:                    295.00          Matches:        54
Percent Similarity:       91.53%          Conservative:   0
Best Local Similarity:    91.53%          Mismatches:     5
Query Match:              90.49%          Indels:         0
DB:                       6               Gaps:           0

US-09-774-954-3 (1-61) x BD103189 (1-1300)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db        142 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 201

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        202 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 261

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 318

RESULT 10
AR163451
LOCUS       AR163451                1514 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 2 from patent US 6270987.
```

```
ACCESSION   AR163451
VERSION     AR163451.1  GI:16234056
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: US 6270987-A 2 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1514
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:                3.84e-32        Length:         1514
Score:                    295.00          Matches:        54
Percent Similarity:       91.53%          Conservative:   0
Best Local Similarity:    91.53%          Mismatches:     5
Query Match:              90.49%          Indels:         0
DB:                       6               Gaps:           0

US-09-774-954-3 (1-61) x AR163451 (1-1514)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db          7 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 66

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db         67 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 126

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 183

RESULT 11
BD103188
LOCUS       BD103188                1514 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103188
VERSION     BD103188.1  GI:22648762
KEYWORDS    JP 2001527389-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 1 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/1
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .1514
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1514
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN
```

```
Alignment Scores:
Pred. No.:                3.84e-32        Length:         1514
Score:                    295.00          Matches:        54
Percent Similarity:       91.53%          Conservative:   0
Best Local Similarity:    91.53%          Mismatches:     5
Query Match:              90.49%          Indels:         0
DB:                       6               Gaps:           0

US-09-774-954-3 (1-61) x BD103188 (1-1514)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db          7 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 66

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db         67 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 126

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 183

RESULT 12
BC000582
LOCUS       BC000582                1524 bp    mRNA    linear   PRI 29-JUN-2004
DEFINITION  Homo sapiens protein O-fucosyltransferase 1, transcript variant 2,
            mRNA (cDNA clone MGC:2482 IMAGE:3162533), complete cds.
ACCESSION   BC000582
VERSION     BC000582.2  GI:37588976
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1524)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1524)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     On Oct 8, 2003 this sequence version replaced gi:12653608.
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
```

```
          Gaithersburg, Maryland;
          Web site: http://www.nisc.nih.gov/
          Contact: nisc_mgc@nhgri.nih.gov
          Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
          Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
          Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
          Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
          Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
          McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
          Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
          Young,A., Zhang,L.-H. and Green,E.D.

          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAL Plate: 5 Row: o Column: 9
          This clone was selected for full length sequencing because it
          passed the following selection criteria: matched mRNA gi: 27436890.
FEATURES             Location/Qualifiers
     source          1. .1524
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:2482 IMAGE:3162533"
                     /tissue_type="Brain, neuroblastoma"
                     /clone_lib="NIH_MGC_19"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1. .1524
                     /gene="POFUT1"
                     /note="synonyms: KIAA0180, O-FUT, O-Fuc-T, FUT12, MGC2482,
                     O-FucT-1"
                     /db_xref="LocusID:23509"
                     /db_xref="MIM:607491"
     CDS             50. .634
                     /gene="POFUT1"
                     /codon_start=1
                     /product="protein O-fucosyltransferase 1, isoform 2
                     precursor"
                     /protein_id="AAH00582.1"
                     /db_xref="GI:12653609"
                     /db_xref="LocusID:23509"
                     /db_xref="MIM:607491"
                     /translation="MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRF
                     GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRV
                     ISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFN
                     KSELFTGISFSASYREQWSQRRENHSCVTLLFPR"
ORIGIN

Alignment Scores:
Pred. No.:              3.87e-32        Length:         1524
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     9               Gaps:           0

US-09-774-954-3 (1-61) x BC000582 (1-1524)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||
Db        125 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 184

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        185 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 244

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 301

RESULT 13
HSM803330
LOCUS       HSM803330              4560 bp    mRNA    linear   PRI 13-MAY-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp451J114 (from clone DKFZp451J114).
ACCESSION   AL832023
VERSION     AL832023.1  GI:21732563
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4560)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp451J114) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1. .4560
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp451J114"
                     /db_xref="taxon:9606"
                     /clone="DKFZp451J114"
                     /tissue_type="human skeletal muscle"
                     /clone_lib="451 (synonym: hlcc1). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
     polyA_site      4538
ORIGIN

Alignment Scores:
Pred. No.:              1.31e-31        Length:         4560
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     9               Gaps:           0

US-09-774-954-3 (1-61) x HSM803330 (1-4560)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||
Db         57 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 116

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        117 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 176

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 233

RESULT 14
D80002
LOCUS       D80002                 5189 bp    mRNA    linear   PRI 10-JAN-2003
DEFINITION  Homo sapiens KIAA0180 mRNA, complete cds.
ACCESSION   D80002
VERSION     D80002.2  GI:20521837
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nagase,T., Seki,N., Ishikawa,K., Tanaka,A. and Nomura,N.
  TITLE     Prediction of the coding sequences of unidentified human genes. V.
```

```
          The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
          analysis of cDNA clones from human cell line KG-1
  JOURNAL   DNA Res. 3 (1), 17-24 (1996)
  MEDLINE   96281124
   PUBMED   8724849
REFERENCE   2
  AUTHORS   Chiang,P.W., Wang,S., Smithivas,P., Song,W.J., Ramamoorthy,S.,
          Hillman,J., Puett,S., Van Keuren,M.L., Crombez,E., Kumar,A.,
          Glover,T.W., Miller,D.E., Tsai,C.H., Blackburn,C.C., Chen,X.N.,
          Sun,Z., Cheng,J.F., Korenberg,J.R. and Kurnit,D.M.
  TITLE     Identification and analysis of the human and murine putative
          chromatin structure regulator SUPT6H and Supt6h
  JOURNAL   Genomics 34 (3), 328-333 (1996)
  MEDLINE   96374824
   PUBMED   8786132
REFERENCE   3  (bases 1 to 5189)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1995) Osamu Ohara, Kazusa DNA Research Institute;
          1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
          (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913)
COMMENT     On May 9, 2002 this sequence version replaced gi:1136419.
FEATURES             Location/Qualifiers
     source          1. .5189
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="ha02567s1"
                     /sex="male"
                     /cell_line="KG-1"
                     /cell_type="myeloblast"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
                     /note="This sequence was obtained by subcloning of the DNA
                     fragments derived from two cDNA clones (1 - 1066 was
                     derived from pf00933 and 1067 - 5189 was derived from
                     ha02567)."
     gene            1. .5189
                     /gene="KIAA0180"
     CDS             <1. .1213
                     /gene="KIAA0180"
                     /citation=[2]
                     /codon_start=2
                     /protein_id="BAA11497.2"
                     /db_xref="GI:20521838"
                     /translation="PPRLCAAAGSGSRADMGAAAWARPLSVSFLLLLLPLPGMPAGSW
                     DPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSY
                     QKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEG
                     NPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE
                     EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHF
                     MASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPEL
                     QQLFKGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFG
                     MDRPPKLRDEF"
ORIGIN

Alignment Scores:
Pred. No.:              1.52e-31        Length:         5189
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     9               Gaps:           0

US-09-774-954-3 (1-61) x D80002 (1-5189)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db        122 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 181

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              |||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        182 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 241
```

```
Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 298

RESULT 15
AF375884
LOCUS       AF375884               5249 bp    mRNA    linear   PRI 23-OCT-2001
DEFINITION  Homo sapiens protein o-fucosyltransferase (POFUT1) mRNA, complete
          cds.
ACCESSION   AF375884
VERSION     AF375884.1  GI:15825113
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5249)
  AUTHORS   Wang,Y., Shao,L., Shi,S., Harris,R.J., Spellman,M.W., Stanley,P.
          and Haltiwanger,R.S.
  TITLE     Modification of epidermal growth factor-like repeats with O-fucose.
          Molecular cloning and expression of a novel GDP-fucose protein
          O-fucosyltransferase
  JOURNAL   J. Biol. Chem. 276 (43), 40338-40345 (2001)
  MEDLINE   21523965
   PUBMED   11524432
REFERENCE   2  (bases 1 to 5249)
  AUTHORS   Shi,S., Stanley,P., Wang,Y., Shao,L., Harris,R.J., Spellman,M.W.
          and Haltiwanger,R.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2001) Cell Biology, Albert Einstein College of
          Medicine, 1300 Morris Park Ave., New York, NY 10461, USA
FEATURES             Location/Qualifiers
     source          1. .5249
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="between PLAGL2 and KIF3B"
                     /tissue_type="heart"
     gene            1. .5249
                     /gene="POFUT1"
     CDS             50. .1216
                     /gene="POFUT1"
                     /function="catalyzes transfer of fucose from GDP-fucose to
                     serine or threonine on epidermal growth factor-like
                     repeats"
                     /note="O-FucT-1; glycosyltransferase"
                     /codon_start=1
                     /product="protein o-fucosyltransferase"
                     /protein_id="AAL09576.1"
                     /db_xref="GI:15825114"
                     /translation="MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRF
                     GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRV
                     ISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFN
                     KSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEM
                     VKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA
                     PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKP
                     EVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF"
ORIGIN

Alignment Scores:
Pred. No.:              1.54e-31        Length:         5249
Score:                  295.00          Matches:        54
Percent Similarity:     91.53%          Conservative:   0
Best Local Similarity:  91.53%          Mismatches:     5
Query Match:            90.49%          Indels:         0
DB:                     9               Gaps:           0

US-09-774-954-3 (1-61) x AF375884 (1-5249)

Qy          3 AlaGlySerTrpAspLeuAlaGlyTyrLeuLeuTyr***Pro***MetGlyArgPheGly 22
              ||||||||||||||   ||||||||||||||||||   |||   ||||||||||||||||
Db        125 GCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGCTTTGGG 184
```

```
Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeu***ValArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||      |||||||||
Db        185 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 244

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 301
```

Search completed: October 26, 2005, 12:55:28
Job time : 1386.21 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:21:43 ; Search time 7.64066 Seconds
(without alignments)
768.157 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGYLLYXPXMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	114	35.0	474	2	T15511	hypothetical prote
2	68.5	21.0	569	2	E86957	probable cholester
3	68.5	21.0	585	2	S72824	cholesterol oxidas
4	67.5	20.7	354	2	I51266	opsin, green-sensi
5	66.5	20.4	767	2	D95101	SpoE family protei
6	66.5	20.4	767	2	E97969	cell division prot
7	63	19.3	562	2	T40967	hypothetical prote
8	62	19.0	578	2	F70736	probable choD prot
9	61.5	18.9	365	2	A36047	galactoside 2-alph
10	61	18.7	674	2	T21217	hypothetical prote
11	59.5	18.3	198	2	C70570	hypothetical prote
12	59.5	18.3	491	2	S76943	hypothetical prote
13	58.5	17.9	219	2	S72721	amidotransferase h
14	58.5	17.9	223	2	B86968	conserved hypothet
15	58.5	17.9	348	2	AE0875	D-erythrose 4-phos
16	57	17.5	363	2	T22255	hypothetical prote
17	56.5	17.3	207	2	A75475	probable acetyltra
18	56.5	17.3	225	2	T21655	hypothetical prote
19	56.5	17.3	227	2	T21662	hypothetical prote
20	56.5	17.3	355	2	A42347	opsin, green-sensi
21	56.5	17.3	467	2	AF0488	amino acid permeas
22	56.5	17.3	492	2	T41004	hypothetical prote
23	56.5	17.3	514	2	T02648	probable phospholi
24	56	17.2	214	2	C69879	hypothetical prote
25	56	17.2	328	2	A70127	phosphate ABC tran
26	55.5	17.0	158	2	AI2377	lipoprotein signal
27	55.5	17.0	231	2	F83032	hypothetical prote
28	55.5	17.0	560	2	T07964	(S)-N-methylcoclau
29	55.5	17.0	1465	2	A70199	hypothetical prote
30	55	16.9	345	2	T12357	NADH2 dehydrogenas
31	55	16.9	428	2	T27763	hypothetical prote
32	54.5	16.7	346	2	T13973	NADH2 dehydrogenas
33	54.5	16.7	383	2	T14795	hypothetical prote
34	54.5	16.7	406	2	AF3342	probable aminotran
35	54.5	16.7	8563	2	T30226	polyketide synthas
36	54	16.6	344	2	T17056	NADH2 dehydrogenas
37	54	16.6	344	2	T17059	NADH2 dehydrogenas
38	54	16.6	344	2	T12335	NADH2 dehydrogenas
39	54	16.6	346	2	B90626	NADH dehydrogenase
40	54	16.6	346	2	T11076	NADH2 dehydrogenas
41	54	16.6	487	1	B64795	ybdS protein - Esc
42	54	16.6	487	2	G85560	probable membrane
43	54	16.6	487	2	C90710	probable membrane
44	54	16.6	501	2	AE0578	citrate carrier [i
45	53.5	16.4	333	1	PARFAS	fructose-bisphosph

ALIGNMENTS

RESULT 1
T15511
hypothetical protein C15C7.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15511
R;Leimbach, D.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid C15C7.
A;Reference number: Z18363
A;Accession: T15511
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-474 <LEI>
A;Cross-references: EMBL:U41528; NID:g1109795; PID:g1109800; PIDN:AAA83156.1; CESP:C15C7.
C;Genetics:
A;Gene: CESP:C15C7.1
A;Introns: 31/3; 67/2; 106/2; 166/1; 202/3; 235/3; 364/3; 410/3

Query Match 35.0%; Score 114; DB 2; Length 474;
Best Local Similarity 64.7%; Pred. No. 3.2e-07;
Matches 22; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

```
Qy         19 GRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQH 52
              |||||| | ||| |||||   ||| :| :||::|
Db        166 GRFGNQVDQFLGVLAFAKALDRTLVLPNFIEFKH 199
```

RESULT 2
E86957
probable cholesterol oxidase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C;Accession: E86957
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squ
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: E86957
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-569 <STO>
A;Cross-references: UNIPROT:Q9CCV1; GB:AL450380; NID:g13092664; PIDN:CAC29897.1; GSPDB:GI
C;Genetics:
A;Gene: choD

Query Match 21.0%; Score 68.5; DB 2; Length 569;
Best Local Similarity 33.8%; Pred. No. 0.45;
Matches 23; Conservative 6; Mismatches 14; Indels 25; Gaps 4;

```
Qy          5 SWDLAGYLLYXPXMGRFGNQADHFL-------------GSLAFAKLXVRTLAVPPWIEYQ 51
              ||||  : |: | :| :| |  | |              ||| :|     || |||
Db         48 SWDLRKF-LWAPKLGCYGIQRIHLLRNVMILAGAGVGGGSLNYA----NTLYVPP----- 97

Qy         52 HHKPPPTN 59
                :| | |
Db         98 --EPFFAN 103
```

RESULT 3
S72824
cholesterol oxidase (EC 1.1.3.6) choD - Mycobacterium leprae
N;Alternate names: B1620_C3_240 protein
C;Species: Mycobacterium leprae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S72824
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A;Description: Mycobacterium leprae cosmid B1620.
A;Reference number: S72584
A;Accession: S72824
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-585 <SMI>
A;Cross-references: UNIPROT:Q59530; EMBL:U00015; NID:g466931; PIDN:AAC43233.1; PID:g4669
C;Genetics:
A;Gene: choD
A;Start codon: GTG
C;Keywords: oxidoreductase

```
  Query Match             21.0%;  Score 68.5;  DB 2;  Length 585;
  Best Local Similarity   33.8%;  Pred. No. 0.47;
  Matches   23;  Conservative    6;  Mismatches   14;  Indels   25;  Gaps    4;

Qy          5 SWDLAGYLLYXPXMGRFGNQADHFL-------------GSLAFAKLXVRTLAVPPWIEYQ 51
              ||||  : |: | :| :| |  | |              ||| :|     || |||
Db         64 SWDLRKF-LWAPKLGCYGIQRIHLLRNVMILAGAGVGGGSLNYA----NTLYVPP----- 113

Qy         52 HHKPPPTN 59
                :| | |
Db        114 --EPFFAN 119
```

RESULT 4
I51266
opsin, green-sensitive - Mexican tetra
C;Species: Astyanax mexicanus (Mexican tetra)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: I51266
R;Register, E.A.; Yokoyama, R.; Yokoyama, S.
J. Mol. Evol. 39, 268-273, 1994
A;Title: Multiple origins of the green-sensitive opsin genes in fish.
A;Reference number: I51266; MUID:95018302; PMID:7932788
A;Accession: I51266
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-354 <REG>
A;Cross-references: UNIPROT:P51474; GB:S75255; NID:g807171; PIDN:AAB32221.1; PID:g807172
C;Genetics:
A;Gene: rh11Af
A;Introns: 124/1; 180/2; 235/3; 315/3
C;Superfamily: vertebrate rhodopsin
C;Keywords: chromoprotein; eye; G protein-coupled receptor; lipoprotein; phosphoprotein;
F;299/Binding site: retinal (Lys) (covalent) #status predicted

```
  Query Match             20.7%;  Score 67.5;  DB 2;  Length 354;
  Best Local Similarity   29.1%;  Pred. No. 0.36;
  Matches   16;  Conservative    6;  Mismatches   28;  Indels    5;  Gaps    1;

Qy          6 WDLA-----GYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKP 55
              | |       |::    || |   | | || : |      | | || : :  : |
Db        129 WSLVVLAIERYIVVCKPMGSFKFSASHALGGIGFTWFMAMTCAAPPLVGWSRYIP 183
```

RESULT 5
D95101
SpoE family protein [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C;Accession: D95101
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heide
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, E
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: D95101
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-767 <KUR>
A;Cross-references: UNIPROT:Q8DQ94; GB:AE005672; PIDN:AAK75005.1; PID:g14972351; GSPDB:GN
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP0878
C;Superfamily: Bacillus subtilis DNA translocase spoIIIE

```
  Query Match             20.4%;  Score 66.5;  DB 2;  Length 767;
  Best Local Similarity   37.2%;  Pred. No. 1.2;
  Matches   16;  Conservative    5;  Mismatches   19;  Indels    3;  Gaps    1;

Qy          8 LAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEY 50
              | |  || |    | |   :|:||:    : | :| | ||  |
Db        140 LIGVALYIPTAFLFSNIGTYFIGSIL---ILVGSLLVSPWSVY 179
```

RESULT 6
E97969
cell division protein [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C;Accession: E97969
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; Es
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; Mc
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: E97969
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-767 <KUR>
A;Cross-references: UNIPROT:Q8DQ94; GB:AE007317; PIDN:AAK99585.1; PID:g15458378; GSPDB:GN
C;Genetics:
A;Gene: ftsK
C;Superfamily: Bacillus subtilis DNA translocase spoIIIE

```
  Query Match             20.4%;  Score 66.5;  DB 2;  Length 767;
  Best Local Similarity   37.2%;  Pred. No. 1.2;
  Matches   16;  Conservative    5;  Mismatches   19;  Indels    3;  Gaps    1;

Qy          8 LAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEY 50
              | |  || |    | |   :|:||:    : | :| | ||  |
Db        140 LIGVALYIPTAFLFSNIGTYFIGSIL---ILVGSLLVSPWSVY 179
```

RESULT 7
T40967
hypothetical protein SPCC1450.18 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T40967; T41000
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, October 1998

```
A;Reference number: Z21961
A;Accession: T40967
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-562 <LYN>
A;Cross-references: UNIPROT:O94575; EMBL:AL031966; PIDN:CAA21436.2; GSPDB:GN00068; SPDB:
A;Experimental source: strain 972h-; cosmid c1442
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, March 1999
A;Reference number: Z21962
A;Accession: T41000
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-562 <LY2>
A;Cross-references: EMBL:AL049559; PIDN:CAB40185.1; GSPDB:GN00068; SPDB:SPCC1450.18
A;Experimental source: strain 972h-; cosmid c1450
C;Genetics:
A;Gene: SPCC1450.18; SPCC1442.02
A;Map position: 3
A;Introns: 58/3
C;Superfamily: Schizosaccharomyces pombe hypothetical protein SPCC1450.18

  Query Match             19.3%;  Score 63;  DB 2;  Length 562;
  Best Local Similarity   44.8%;  Pred. No. 2.4;
  Matches   13;  Conservative    6;  Mismatches    8;  Indels    2;  Gaps    1;

Qy          5 SWDLAGYLLYXPXMGRFGNQADHFLGSLA 33
              |||: | :|  .| :|:  |:|| |:  ||
Db         66 SWDIIGMVL--PYVGKVQNEADEFVDFLA 92


RESULT 8
F70736
probable choD protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C;Accession: F70736
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: F70736
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-578 <COL>
A;Cross-references: UNIPROT:Q57307; GB:Z77165; GB:AL123456; NID:g3261609; PIDN:CAB01014.
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: choD

  Query Match             19.0%;  Score 62;  DB 2;  Length 578;
  Best Local Similarity   36.4%;  Pred. No. 3.4;
  Matches   20;  Conservative    5;  Mismatches   12;  Indels   18;  Gaps    3;

Qy          5 SWDLAGYLLYXPXMGRFGNQADHFL----------GSLAFAKLXVRTLAVPP 46
              ||||  : |: | :| :| | | |             ||| :|     || |||
Db         48 SWDLRKF-LWAPRLGCYGIQRIHPLRNVMILAGAGVGGGSLNYA----NTLYVPP 97


RESULT 9
A36047
galactoside 2-alpha-L-fucosyltransferase (EC 2.4.1.69) - human
C;Species: Homo sapiens (man)
C;Date: 16-Nov-1990 #sequence_revision 16-Nov-1990 #text_change 09-Jul-2004
C;Accession: A36047
R;Larsen, R.D.; Ernst, L.K.; Nair, R.P.; Lowe, J.B.
Proc. Natl. Acad. Sci. U.S.A. 87, 6674-6678, 1990
A;Title: Molecular cloning, sequence, and expression of a human GDP-L-fucose:beta-D-gala
A;Reference number: A36047; MUID:90370848; PMID:2118655
A;Accession: A36047
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-365 <LAR>
A;Cross-references: UNIPROT:P19526; GB:M35531; NID:g183887; PID:g306830
C;Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

  Query Match             18.9%;  Score 61.5;  DB 2;  Length 365;
  Best Local Similarity   27.4%;  Pred. No. 2.4;
  Matches   26;  Conservative    6;  Mismatches   20;  Indels   43;  Gaps    5;

Qy          2 LAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVR----------------TLA 43
              |:|:| :       |  ||||||   :   || |:|  |                  ||
Db         77 LSGTWTV------YPN-GRFGNQMGQYATLLALAQLNGRRAFILPAMHAALAPVFRITLP 129

Qy         44 V--------PPWIEYQHH----------KPPFTNL 60
              |        || | | |          : ||  |
Db        130 VLAPEVDSRTPWRELQLHDWMSEEYADLRDPFLKL 164


RESULT 10
T21217
hypothetical protein F21G4.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T21217
R;Mortimore, B.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z19392
A;Accession: T21217
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-674 <WIL>
A;Cross-references: UNIPROT:Q93550; EMBL:Z81016; PIDN:CAB02665.1; GSPDB:GN00028; CESP:F21
A;Experimental source: clone F21G4
C;Genetics:
A;Gene: CESP:F21G4.1
A;Map position: X
A;Introns: 58/3; 149/1; 181/3; 213/3; 286/1; 327/2; 457/2; 509/3; 608/1

  Query Match             18.7%;  Score 61;  DB 2;  Length 674;
  Best Local Similarity   27.4%;  Pred. No. 5.4;
  Matches   17;  Conservative   14;  Mismatches   25;  Indels    6;  Gaps    2;

Qy          4 GSWDLAGYLLYXPXM-----GRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFT 58
              ||: : ||  | |       ||  : || : | ::   : | | |:  ||  :::  | :
Db        359 GSYLIGGYQTYLPKFIETQYGRSASMADIYSGIISVGAIAVST-ALGGWILSRYNIAPRS 417

Qy         59 NL 60
              ::
Db        418 SI 419


RESULT 11
C70570
hypothetical protein Rv2604c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C;Accession: C70570
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: C70570
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-198 <COL>
A;Cross-references: UNIPROT:O06210; GB:Z95387; GB:AL123456; NID:g3261763; PIDN:CAB08636.1
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv2604c
```

```
C;Superfamily: conserved hypothetical protein HI1648

  Query Match             18.3%;  Score 59.5;  DB 2;  Length 198;
  Best Local Similarity   31.8%;  Pred. No. 2.2;
  Matches   21;  Conservative    6;  Mismatches   20;  Indels   19;  Gaps    3;

Qy          3 AGSWDLAGYLLYXPXMGR----------------FGNQADHFLGSLAFAKLX--VRTLAV 44
              ||    ||  :|     ||                ||:| | | | : || |   || : :
Db         82 AGMILLASEILDAGAAGRQALPLRAMNMTVRRNAFGSQVDSFEGDIEFAGLDDPVRAVFI 141

Qy         45 -PPWIE 49
                ||:|
Db        142 RAPWVE 147


RESULT 12
S76943
hypothetical protein - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S76943
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S76943
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-491 <KAN>
A;Cross-references: UNIPROT:P74735; EMBL:D90917; GB:AB001339; NID:g1653836; PIDN:BAA1885
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

  Query Match             18.3%;  Score 59.5;  DB 2;  Length 491;
  Best Local Similarity   26.6%;  Pred. No. 6.1;
  Matches   17;  Conservative    9;  Mismatches   27;  Indels   11;  Gaps    3;

Qy          6 WDLAGYLLYXPXMGR---------FGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPP 56
              |: :||:||     :          |:|    |  :   |    | : | || |:    ||
Db         90 WN-SGYVLYESEFAKREGLQPFSPRGDQGQDVLADI-IDKAHRRNMLVLPWFEFGFKAPP 147

Qy         57 FTNL 60
                : |
Db        148 MSEL 151


RESULT 13
S72721
amidotransferase hisH homolog - Mycobacterium leprae
N;Alternate names: B1177_C1_149
C;Species: Mycobacterium leprae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S72721
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A;Description: Mycobacterium leprae cosmid B1177.
A;Reference number: S72694
A;Accession: S72721
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-219 <SMI>
A;Cross-references: UNIPROT:Q49637; EMBL:U00011; NID:g466807; PIDN:AAA17085.1; PID:g4668
C;Genetics:
A;Gene: hisH
C;Superfamily: conserved hypothetical protein HI1648

  Query Match             17.9%;  Score 58.5;  DB 2;  Length 219;
  Best Local Similarity   43.8%;  Pred. No. 3.4;
  Matches   14;  Conservative    4;  Mismatches   11;  Indels    3;  Gaps    2;

Qy         21 FGNQADHFLGSLAFAKLX--VRTLAV-PPWIE 49
                || | | | | : || |   || : :  ||:|
Db        137 FGRQVDSFEGDIGFAGLVDPVRAVFIRAPWVE 168


RESULT 14
B86968
conserved hypothetical protein ML0474 [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C;Accession: B86968
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squ
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: B86968
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-223 <STO>
A;Cross-references: UNIPROT:Q9CCT5; GB:AL450380; NID:g13092704; PIDN:CAC29982.1; GSPDB:GN
C;Genetics:
A;Gene: ML0474
C;Superfamily: conserved hypothetical protein HI1648

  Query Match             17.9%;  Score 58.5;  DB 2;  Length 223;
  Best Local Similarity   43.8%;  Pred. No. 3.4;
  Matches   14;  Conservative    4;  Mismatches   11;  Indels    3;  Gaps    2;

Qy         21 FGNQADHFLGSLAFAKLX--VRTLAV-PPWIE 49
                || | | | | : || |   || : :  ||:|
Db        141 FGRQVDSFEGDIGFAGLVDPVRAVFIRAPWVE 172


RESULT 15
AE0875
D-erythrose 4-phosphate dehydrogenase [imported] - Salmonella enterica subsp. enterica se
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AE0875
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serova
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AE0875
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-348 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD02901.1; PID:g16504154; GSPDB:GN00176
C;Genetics:
A;Gene: STY3228
C;Superfamily: glyceraldehyde-3-phosphate dehydrogenase

  Query Match             17.9%;  Score 58.5;  DB 2;  Length 348;
  Best Local Similarity   35.8%;  Pred. No. 5.6;
  Matches   19;  Conservative    4;  Mismatches   25;  Indels    5;  Gaps    1;

Qy          2 LAGSWDLAGYLLYXPXMGRFGNQADH-----FLGSLAFAKLXVRTLAVPPWIE 49
              || :  :|  | |     |||  :  |     |:|       |  ||||  || |
Db         38 LADAAGMAHLLKYDTSHGRFAWEVRHEREQLFVGDDVIRILHERTLADLPWRE 90


Search completed: October 25, 2005, 15:36:43
Job time : 9.64066 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:20:23 ; Search time 38.8296 Seconds
(without alignments)
607.588 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGYLLYXPXMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	320	98.2	61	2	AAW80572	Aaw80572 N-termina
2	295	90.5	61	2	AAW80574	Aaw80574 N-termina
3	295	90.5	365	2	AAW80571	Aaw80571 Human hea
4	295	90.5	388	8	ABM82223	Abm82223 Tumour-as
5	295	90.5	397	2	AAW80573	Aaw80573 Human O-f
6	232.5	71.3	417	4	AAM78377	Aam78377 Human pro
7	220.5	67.6	417	4	AAM79361	Aam79361 Human pro
8	204	62.6	343	2	AAW80577	Aaw80577 Partial h
9	177	54.3	402	4	ABB63790	Abb63790 Drosophil
10	81	24.8	28	2	AAW80575	Aaw80575 N-termina
11	76	23.3	150	3	AAY87325	Aay87325 Human sig
12	69.5	21.3	198	6	ABU33866	Abu33866 Protein e
13	66.5	20.4	737	8	ADK48560	Adk48560 Streptoco
14	66.5	20.4	741	6	ABU01274	Abu01274 S. pneumo
15	66.5	20.4	763	8	ADR94975	Adr94975 Novel S.
16	66.5	20.4	767	4	AAU37750	Aau37750 Streptoco
17	66.5	20.4	767	4	AAU38005	Aau38005 Streptoco
18	66.5	20.4	767	6	ABU45987	Abu45987 Protein e
19	66.5	20.4	768	3	AAY81596	Aay81596 Streptoco
20	66.5	20.4	783	2	AAW97416	Aaw97416 A membran
21	61.5	18.9	365	2	AAR13751	Aar13751 GDP-Fuc:b
22	61.5	18.9	365	2	AAR45936	Aar45936 A glycosy
23	61.5	18.9	365	2	AAR70421	Aar70421 2-Alpha-f
24	61.5	18.9	365	2	AAR80154	Aar80154 GDP-L-fuc
25	61.5	18.9	365	2	AAR70422	Aar70422 2-Alpha-f
26	61.5	18.9	365	2	AAR90572	Aar90572 Human H-t
27	61.5	18.9	365	2	AAW23805	Aaw23805 Human alp
28	61.5	18.9	365	2	AAW13640	Aaw13640 Human alp
29	61.5	18.9	365	3	AAY97279	Aay97279 Human H-t
30	61.5	18.9	365	5	AAE29230	Aae29230 Human alp
31	60.5	18.6	580	8	ADN25627	Adn25627 Bacterial
32	59.5	18.3	198	6	ABU36823	Abu36823 Protein e
33	59.5	18.3	198	6	ABU34438	Abu34438 Protein e
34	59.5	18.3	340	2	AAR63066	Aar63066 Acetyl-es
35	59	18.1	366	6	ABM69163	Abm69163 Photorhab
36	58.5	17.9	223	6	ABU35781	Abu35781 Protein e
37	58.5	17.9	365	2	AAW53102	Aaw53102 Pig H tra
38	58.5	17.9	365	2	AAW30630	Aaw30630 Swine alp
39	58.5	17.9	365	2	AAW97356	Aaw97356 Swine alp
40	58.5	17.9	365	3	AAY79302	Aay79302 Pig alpha
41	58.5	17.9	365	5	AAB47995	Aab47995 Swine alp
42	57	17.5	322	8	ADR88379	Adr88379 Caenorhab
43	57	17.5	363	8	ADN23009	Adn23009 Bacterial
44	56.5	17.3	514	5	ABB91896	Abb91896 Herbicida
45	56	17.2	245	5	ABP42996	Abp42996 Human ova

ALIGNMENTS

RESULT 1
AAW80572
ID AAW80572 standard; peptide; 61 AA.
XX
AC AAW80572;
XX
DT 16-DEC-1998 (first entry)
XX
DE N-terminal amino acid sequence of CHO O-fucosyltransferase.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 15
FT /note= "unknown"
FT Misc-difference 17
FT /note= "unknown"
FT Misc-difference 38
FT /note= "unknown"
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Claim 5; Page 69; 90pp; English.
XX
CC This represents the N-terminal sequence of CHO O-fucosyltransferase. The
CC enzyme can glycosylate an epidermal growth factor (EGF) domain of a
CC polypeptide with an activated O-fucose residue. Inhibitors of O-
CC fucosyltransferase, e.g. mutants with increased affinity for the EFG

```
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 61 AA;

  Query Match             98.2%;  Score 320;  DB 2;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 9.3e-38;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RLAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RLAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 2
AAW80574
ID   AAW80574 standard; peptide; 61 AA.
XX
AC   AAW80574;
XX
DT   16-DEC-1998  (first entry)
XX
DE   N-terminal amino acid sequence of human heart O-fucosyltransferase.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS   Homo sapiens.
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Claim 4; Page 69; 90pp; English.
XX
CC   This represents a the N-terminal sequence of the human heart O-
CC   fucosyltransferase that can glycosylate an epidermal growth factor (EGF)
CC   domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC   O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 61 AA;

  Query Match             90.5%;  Score 295;  DB 2;  Length 61;
  Best Local Similarity   91.5%;  Pred. No. 3.5e-34;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              |||||  ||||||  | ||||||||||||||||||||  ||||||||||||||||||||
Db          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 3
AAW80571
ID   AAW80571 standard; protein; 365 AA.
XX
AC   AAW80571;
XX
DT   16-DEC-1998  (first entry)
XX
DE   Human heart O-fucosyltransferase.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .61
FT                   /note= "N-terminal sequence claimed for in claim 4"
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
DR   N-PSDB; AAV65632.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example 1; Fig 12A; 90pp; English.
XX
CC   This represents a human heart O-fucosyltransferase that can glycosylate
CC   an epidermal growth factor (EGF) domain of a polypeptide with an
CC   activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.
CC   mutants with increased affinity for the EFG domains, are used in
CC   diagnosis and treatment of conditions associated with overexpression of O
CC   -fucosyltransferase, to promote survival of sensory (retinal) neurons.
CC   Probes based on EGF domain polypeptide are used to detect gene
CC   amplification and expression. The expression can also be determined at
CC   the protein level using antibodies specific for O-fucosyltransferase
XX
SQ   Sequence 365 AA;

  Query Match             90.5%;  Score 295;  DB 2;  Length 365;
  Best Local Similarity   91.5%;  Pred. No. 2.9e-33;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              |||||  ||||||  | ||||||||||||||||||||  ||||||||||||||||||||
Db          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 4
ABM82223
ID   ABM82223 standard; protein; 388 AA.
XX
```

```
AC   ABM82223;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Tumour-associated antigenic target (TAT) polypeptide PRO83146, SEQ:5726.
XX
KW   Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW   tumour; diagnosis; cell proliferative disorder; breast cancer;
KW   colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW   central nervous system cancer; bladder cancer; pancreatic cancer;
KW   cervical cancer; melanoma; leukaemia; hybridisation probe;
KW   chromosome identification; chromosome mapping; gene mapping;
KW   gene therapy; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004030615-A2.
XX
PD   15-APR-2004.
XX
PF   29-SEP-2003; 2003WO-US028547.
XX
PR   02-OCT-2002; 2002US-0414971P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wu TD,  Zhang Z,  Zhou Y;
XX
DR   WPI; 2004-347921/32.
DR   N-PSDB; ACN40746.
XX
PT   New tumor-associated antigenic target polypeptides and nucleic acids,
PT   useful in preparing a medicament for treating or detecting a
PT   proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT   prostate cancer or tumor.
XX
PS   Claim 12; SEQ ID NO 5726; 7273pp; English.
XX
CC   The invention relates to human tumour-associated antigenic target (TAT)
CC   polypeptides, and their related nucleic acids. The TAT polypeptides are
CC   overexpressed in cancer tissues compared to normal tissues, and may thus
CC   serve as effective targets for the diagnosis and treatment of cancer in
CC   mammals. The invention also relates to nucleic acid and polypeptide
CC   sequences at least 80% identical to the TAT nucleic acids and
CC   polypeptides; expression vectors and host cells comprising a TAT nucleic
CC   acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC   molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC   TAT polypeptide; and methods and compositions for the treatment or
CC   diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC   antibodies, antagonists, binding molecules and compositions are useful
CC   for diagnosing or treating a cell proliferative disorder associated with
CC   increased TAT expression, particularly cancers such as breast cancer,
CC   colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC   cancer, pancreatic cancer, cervical cancer, cancers of the central
CC   nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC   used as hybridisation probes, in chromosome and gene mapping, in
CC   chromosome identification and in gene therapy. The present sequence
CC   represents a TAT polypeptide of the invention
XX
SQ   Sequence 388 AA;

  Query Match             90.5%;  Score 295;  DB 8;  Length 388;
  Best Local Similarity   91.5%;  Pred. No. 3.1e-33;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | |||||||||||||||||||  ||||||||||||||||||||||||
Db         26 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 84


RESULT 5
AAW80573
ID   AAW80573 standard; protein; 397 AA.
XX
AC   AAW80573;
XX
DT   16-DEC-1998  (first entry)
XX
DE   Human O-fucosyltransferase sequence expressed by a plasmid insertion.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         27. .32
FT                   /note= "polyhistidine tag"
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
DR   N-PSDB; AAV65633.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example; Fig 13B; 90pp; English.
XX
CC   This represents the amino acid sequence of the human heart O-
CC   fucosyltransferase expressed by a plasmid insertion. The human O-
CC   fucosyltransferase can glycosylate an epidermal growth factor (EGF)
CC   domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC   O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 397 AA;

  Query Match             90.5%;  Score 295;  DB 2;  Length 397;
  Best Local Similarity   91.5%;  Pred. No. 3.2e-33;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | |||||||||||||||||||  ||||||||||||||||||||||||
Db         35 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 93


RESULT 6
AAM78377
ID   AAM78377 standard; protein; 417 AA.
XX
AC   AAM78377;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 1039.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
```

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK51510.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 20; Page 3276-3277; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 417 AA;

Query Match 71.3%; Score 232.5; DB 4; Length 417;
Best Local Similarity 73.3%; Pred. No. 3e-24;
Matches 44; Conservative 4; Mismatches 7; Indels 5; Gaps 1;

```
Qy          7 DLAGYLLYXPXM-----GRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              |:| :  : | :     |||||||||||||||||  |||||||||||||||||||||
Db         54 DMATWPTHAPSVEEGGQGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 113
```

RESULT 7
AAM79361
ID AAM79361 standard; protein; 417 AA.
XX
AC AAM79361;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3007.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR N-PSDB; AAK52494.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 20; Page 229-230; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 417 AA;

Query Match 67.6%; Score 220.5; DB 4; Length 417;
Best Local Similarity 70.0%; Pred. No. 1.5e-22;
Matches 42; Conservative 5; Mismatches 8; Indels 5; Gaps 1;

```
Qy          7 DLAGYLLYXPXM-----GRFGNQADHFLGSLAFAKLLVRTLAVPPWIEYQHHKPPFTNLH 61
              |:| :  : | :     |||||||||||||||||  |:|||| ||||||||||||||
Db         54 DMATWPTHAPSVEEGGQGRFGNQADHFLGSLAFAKLLNRSLAVPSWIEYQHHKPPFTNLH 113
```

RESULT 8
AAW80577
ID AAW80577 standard; protein; 343 AA.
XX
AC AAW80577;
XX
DT 16-DEC-1998 (first entry)
XX
DE Partial human sequence of unknown function from a myeblast cell line.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS Homo sapiens.
XX

```
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
DR   N-PSDB; AAV65634.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example 1; Page 62-68; 90pp; English.
XX
CC   This represents a published partial human sequence of unknown function
CC   from a myeblast cell line. The invention provides a human heart O-
CC   fucosyltransferase enzyme that can glycosylate an epidermal growth factor
CC   (EGF) domain of a polypeptide with an activated O-fucose residue.
CC   Inhibitors of O-fucosyltransferase, e.g. mutants with increased affinity
CC   for the EFG domains, are used in diagnosis and treatment of conditions
CC   associated with overexpression of O-fucosyltransferase, to promote
CC   survival of sensory (retinal) neurons. Probes based on EGF domain
CC   polypeptide are used to detect gene amplification and expression. The
CC   expression can also be determined at the protein level using antibodies
CC   specific for O-fucosyltransferase
XX
SQ   Sequence 343 AA;

  Query Match             62.6%;  Score 204;  DB 2;  Length 343;
  Best Local Similarity   94.9%;  Pred. No. 2.8e-20;
  Matches   37;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         23 NQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              |||||||||||||||  ||||||||||||||||||||||
Db          1 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 39


RESULT 9
ABB63790
ID   ABB63790 standard; protein; 402 AA.
XX
AC   ABB63790;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 18162.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL07893.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 18162; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 402 AA;

  Query Match             54.3%;  Score 177;  DB 4;  Length 402;
  Best Local Similarity   73.3%;  Pred. No. 2.5e-16;
  Matches   33;  Conservative    3;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          7 DLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQ 51
              |  ||| | | ||||||||||||||||||||   ||| :|||:||:
Db         28 DPNGYLTYCPCMGRFGNQADHFLGSLAFAKALNRTLILPPWVEYR 72


RESULT 10
AAW80575
ID   AAW80575 standard; protein; 28 AA.
XX
AC   AAW80575;
XX
DT   16-DEC-1998  (first entry)
XX
DE   N-terminal sequence of expressed human O-fucosyltransferase.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 23
FT                   /note= "unknown"
FT   Misc-difference 25
FT                   /note= "unknown"
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example 1; Page 39; 90pp; English.
```

```
XX
CC   This represents the N-terminal amino acid sequence of the human heart O-
CC   fucosyltransferase expressed by a plasmid insertion. The human O-
CC   fucosyltransferase can glycosylate an epidermal growth factor (EGF)
CC   domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC   O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 28 AA;

  Query Match             24.8%;  Score 81;  DB 2;  Length 28;
  Best Local Similarity   94.4%;  Pred. No. 0.00057;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGR 20
              ||||| ||||||||||||
Db         11 AGSWDPAGYLLYXPXMGR 28


RESULT 11
AAY87325
ID   AAY87325 standard; protein; 150 AA.
XX
AC   AAY87325;
XX
DT   11-MAY-2000  (first entry)
XX
DE   Human signal peptide containing protein HSPP-102 SEQ ID NO:102.
XX
KW   Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
KW   inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW   antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;
KW   antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW   reproductive disorder; developmental disorder; arteriosclerosis;
KW   cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW   asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW   Parkinson's disease; Huntington's diseases; ovulatory defect;
KW   muscular dystrophy.
XX
OS   Homo sapiens.
XX
PN   WO200000610-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US014484.
XX
PR   26-JUN-1998;   98US-0090762P.
PR   31-JUL-1998;   98US-0094983P.
PR   01-OCT-1998;   98US-0102686P.
PR   11-DEC-1998;   98US-0112129P.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Lal P,  Tang YT,  Gorgone GA,  Corley NC,  Guegler KJ,  Baughn MR;
PI   Akerblom IE,  Au-Young J,  Yue H,  Patterson C,  Reddy R,  Hillman JL;
PI   Bandman O;
XX
DR   WPI; 2000-160673/14.
DR   N-PSDB; AAZ98210.
XX
PT   New human signal peptide-containing proteins useful in treatment,
PT   prevention and diagnosis of e.g. cancer, inflammation and cardiovascular
PT   disease.
XX
PS   Claim 1; Page 225-226; 327pp; English.
XX
CC   AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the
CC   human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
CC   anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,
CC   neuroprotective, cardiovascular and antiasthmatic activities, and can be
CC   used in gene therapy. HSPPs can be used to treat or prevent disorders
CC   associated with decreased activity or function of HSPP. Antagonists of
CC   HSPP are used to treat or prevent disorders associated with increased
CC   activity or function of HSPP. Such diseases include cell proliferation
CC   (including cancer), inflammation, cardiovascular, neurological,
CC   reproductive or developmental disorders, (e.g. arteriosclerosis,
CC   cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
CC   asthma, Crohn's disease, microbial or other infections, congestive or
CC   ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
CC   diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
CC   nucleic acids can be used for the recombinant production of HSPP, for
CC   detecting HSPP in standard hybridisation and amplification assays (for
CC   diagnosis and monitoring), in gene therapy, as antisense, triplex-forming
CC   or ribozyme therapeutics, for detecting related sequences or genetic
CC   variations, and for chromosomal mapping. HSPP are also used to raise
CC   specific antibodies (Ab) and to screen for agonists and antagonists
CC   (potential therapeutic agents). Ab are used to diagnose, or monitor, HSPP
CC   -related diseases (in usual immunoassays), as therapeutic antagonists, in
CC   competitive drug screens, and for purification of HSPP from natural
CC   sources
XX
SQ   Sequence 150 AA;

  Query Match             23.3%;  Score 76;  DB 3;  Length 150;
  Best Local Similarity   77.8%;  Pred. No. 0.022;
  Matches   14;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGR 20
              ||||| |||||| | ||:
Db         26 AGSWDPAGYLLYCPCMGK 43


RESULT 12
ABU33866
ID   ABU33866 standard; protein; 198 AA.
XX
AC   ABU33866;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #19393.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Mycobacterium avium.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA37736.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
```

PS Claim 25; SEQ ID NO 61790; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 198 AA;

Query Match 21.3%; Score 69.5; DB 6; Length 198;
Best Local Similarity 50.0%; Pred. No. 0.25;
Matches 16; Conservative 2; Mismatches 11; Indels 3; Gaps 1;

```
Qy         21 FGNQADHFLGSLAFAKLXVRTLAV---PPWIE 49
              || | | | | :||| | |   ||    ||:|
Db        116 FGRQVDSFEGDIAFAGLDVPVRAVFIRAPWVE 147
```

RESULT 13
ADK48560
ID ADK48560 standard; protein; 737 AA.
XX
AC ADK48560;
XX
DT 20-MAY-2004 (first entry)
XX
DE Streptococcus pneumoniae protein, Seq ID No 5075.
XX
KW Antibacterial; Gene therapy; Vaccine; Streptococcus pneumoniae.
XX
OS Streptococcus pneumoniae.
XX
PN US6699703-B1.
XX
PD 02-MAR-2004.
XX
PF 26-MAY-2000; 2000US-00583110.
XX
PR 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
PR 30-JUN-1998; 98US-00107433.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm L, Bush D, Zeng Q, Opperman T, Houseweart CE;
XX
DR WPI; 2004-212399/20.
DR N-PSDB; ADK45899.
XX
PT New nucleic acid molecules and polypeptides useful for diagnosing,
PT preventing and treating pathological conditions resulting from bacterial
PT infection, e.g. Streptococcus pneumoniae infection, and in drug
PT screening.
XX
PS Disclosure; SEQ ID NO 5075; 301pp; English.
XX
CC The invention relates to isolated Streptococcus pneumoniae nucleic acids
CC and polypeptides. The nucleic acids and proteins are useful for
CC diagnosing, preventing and treating pathological conditions resulting
CC from bacterial infection, such as S. pneumoniae infection. These may also
CC be used for drug screening procedures. The present sequence represents a
CC Streptococcus pneumoniae polypeptide of the invention. Note: The sequence
CC data for this patent did not appear in the printed specification but was
CC obtained in electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 737 AA;

Query Match 20.4%; Score 66.5; DB 8; Length 737;
Best Local Similarity 37.2%; Pred. No. 3.2;
Matches 16; Conservative 5; Mismatches 19; Indels 3; Gaps 1;

```
Qy          8 LAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEY 50
              | |  || |    | |   :|:||:    : | :| | ||  |
Db        140 LIGVALYIPTAFLFSNIGTYFIGSIL---ILVGSLLVSPWSVY 179
```

RESULT 14
ABU01274
ID ABU01274 standard; protein; 741 AA.
XX
AC ABU01274;
XX
DT 23-OCT-2003 (revised)
DT 11-FEB-2003 (first entry)
XX
DE S. pneumoniae type 4 strain protein from coding region #849.
XX
KW Bacterial meningitis; pneumonia; sepsis; otitis media; ear infection;
KW antiinflammatory; antibacterial; immunostimulant; auditory; respiratory;
KW gene therapy; vaccine.
XX
OS Streptococcus pneumoniae; type 4 strain.
XX
PN WO200277021-A2.
XX
PD 03-OCT-2002.
XX
PF 27-MAR-2002; 2002WO-IB002163.
XX
PR 27-MAR-2001; 2001GB-00007658.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Masignani V, Tettelin H, Fraser C;
XX
DR WPI; 2003-040579/03.
DR N-PSDB; ABX06561.
XX
PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media or
PT ear infection.
XX
PS Claim 1; SEQ ID NO 1698; 56pp; English.
XX
CC The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the

CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC ABS56454. Also included are an antibody which binds one of the proteins,
CC treating a patient by administering the protein, DNA or antibody (in a
CC composition), a kit comprising first and second primers, which are the
CC nucleic acid cited above or fragments between nucleotides 8-100 of a
CC sequence not defined in the specification, for amplifying a target
CC sequence contained within a Streptococcus nucleic acid sequence, where
CC the first primer is substantially complementary to the target sequence
CC and the second primer is substantially complementary to the complement of
CC the target sequence, and where the parts of the primers having
CC substantial complementarity define the termini of the target sequence to
CC be amplified, assay comprising contacting a test compound with the
CC protein, and determining whether the test compound binds to the protein
CC and a Streptococcus pneumoniae bacterium, where one or more genes
CC encoding the proteins has been rendered inactive. The proteins, nucleic
CC acid molecules, antibody and compositions are useful as medicaments for
CC treating or preventing a disease or infection due to streptococcus
CC bacteria, particularly S. pneumoniae, such as pneumonia, sepsis, otitis
CC media or ear infection. They are also useful in developing vaccines,
CC diagnostics and antibiotics. The methods are useful for identifying
CC immunodominant proteins. The present sequence is one of the 2469 proteins
CC expressed by the identified coding regions from the genomic sequence.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences. (Updated on 23-OCT-2003 to
CC standardise OS field)
XX
SQ Sequence 741 AA;

Query Match 20.4%; Score 66.5; DB 6; Length 741;
Best Local Similarity 37.2%; Pred. No. 3.2;
Matches 16; Conservative 5; Mismatches 19; Indels 3; Gaps 1;

Qy 8 LAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEY 50
Db 114 LIGVALYIPTAFLFSNIGTYFIGSIL---ILVGSLLVSPWSVY 153

RESULT 15
ADR94975
ID ADR94975 standard; protein; 763 AA.
XX
AC ADR94975;
XX
DT 16-DEC-2004 (first entry)
XX
DE Novel S. pneumoniae protein sequence, SEQ ID 3610.
XX
KW Meningitis; bacteraemia; pneumonia; otitis media; vaccine;
KW bacterial infection.
XX
OS Streptococcus pneumoniae.
XX
PN US6800744-B1.
XX
PD 05-OCT-2004.
XX
PF 30-JUN-1998; 98US-00107433.
XX
PR 02-JUL-1997; 97US-0051553P.
PR 12-MAY-1998; 98US-0085131P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
DR WPI; 2004-697205/68.
DR N-PSDB; ADR92372.
XX
PT New isolated nucleic acid encoding a Streptococcus pneumoniae
PT polypeptide, useful for diagnosing, preventing and/or treating
PT pathological conditions resulting from the bacterial infection.
XX
PS Disclosure; SEQ ID NO 3610; 151pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a sequence
CC encoding a Streptococcus pneumoniae ADR91366polypeptide, or its
CC fragments, with any of 9 fully defined sequences (appearing as ADR94308,
CC ADR94489, ADR94800, ADR94837, ADR94969, ADR95253, ADR95642, ADR95682,
CC ADR96079) or any of the fully defined sequences appearing as ADR91705,
CC ADR91886, ADR92197, ADR92234, ADR93039, ADR93079, ADR92366, ADR92650 or
CC ADR93476 or at least 20 or 30 consecutive nucleotides of the nucleotide
CC sequences, or at least 40, 60 or 300 consecutive nucleotides, which is
CC hybridisable under high stringency conditions to the nucleotide sequence.
CC The nucleic acids and proteins are chosen from 5206 disclosed sequences.
CC Also included are a recombinant expression vector comprising the isolated
CC nucleic acid cited above operably linked to a transcription regulatory
CC element, a cell comprising the recombinant expression vector and a probe
CC comprising at least 20 consecutive nucleotides of the nucleotide
CC sequences as cited above. The methods and compositions of the present
CC invention are useful for the diagnosis, prevention and/or treatment of
CC pathological conditions resulting from bacterial infection by
CC Streptococcus pneumoniae e.g. pneumonia, bacteraemia, meningitis and
CC otitis media. The present sequence is one of the 2603 disclosed S.
CC pneumoniae protein sequences. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=6800744B1.
XX
SQ Sequence 763 AA;

Query Match 20.4%; Score 66.5; DB 8; Length 763;
Best Local Similarity 37.2%; Pred. No. 3.4;
Matches 16; Conservative 5; Mismatches 19; Indels 3; Gaps 1;

Qy 8 LAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEY 50
Db 166 LIGVALYIPTAFLFSNIGTYFIGSIL---ILVGSLLVSPWSVY 205

Search completed: October 25, 2005, 15:30:40
Job time : 41.8296 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:25:36 ; Search time 10.0205 Seconds
(without alignments)
454.426 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGYLLYXPXMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	320	98.2	61	3	US-08-978-741-3	Sequence 3, Appli
2	320	98.2	61	3	US-09-333-729A-5	Sequence 5, Appli
3	295	90.5	61	3	US-08-978-741-9	Sequence 9, Appli
4	295	90.5	61	3	US-09-333-729A-4	Sequence 4, Appli
5	295	90.5	365	3	US-08-978-741-2	Sequence 2, Appli
6	295	90.5	365	3	US-09-333-729A-3	Sequence 3, Appli
7	295	90.5	397	3	US-08-978-741-6	Sequence 6, Appli
8	295	90.5	397	3	US-09-333-729A-7	Sequence 7, Appli
9	204	62.6	343	3	US-08-978-741-17	Sequence 17, Appl
10	204	62.6	343	3	US-09-333-729A-13	Sequence 13, Appl
11	177	54.3	399	4	US-09-270-767-45921	Sequence 45921, A
12	114	35.0	474	3	US-08-978-741-8	Sequence 8, Appli
13	114	35.0	474	3	US-09-333-729A-12	Sequence 12, Appl
14	81	24.8	28	3	US-08-978-741-14	Sequence 14, Appl
15	81	24.8	28	3	US-09-333-729A-14	Sequence 14, Appl
16	66.5	20.4	737	4	US-09-583-110-5075	Sequence 5075, Ap
17	66.5	20.4	763	4	US-09-107-433-3610	Sequence 3610, Ap
18	66.5	20.4	783	2	US-08-922-837-2	Sequence 2, Appli
19	66.5	20.4	783	3	US-09-351-550-2	Sequence 2, Appli
20	61.5	18.9	333	5	PCT-US91-00899-11	Sequence 11, Appl
21	61.5	18.9	365	1	US-07-914-281-6	Sequence 6, Appli
22	61.5	18.9	365	1	US-08-393-246-6	Sequence 6, Appli
23	61.5	18.9	365	1	US-08-273-411-1	Sequence 1, Appli
24	61.5	18.9	365	1	US-08-525-058A-6	Sequence 6, Appli
25	61.5	18.9	365	1	US-08-395-800A-6	Sequence 6, Appli
26	61.5	18.9	365	1	US-08-395-800A-10	Sequence 10, Appl
27	61.5	18.9	365	2	US-08-696-731-6	Sequence 6, Appli
28	61.5	18.9	365	3	US-09-042-531-6	Sequence 6, Appli
29	61.5	18.9	365	3	US-09-254-077A-11	Sequence 11, Appl
30	61.5	18.9	365	5	PCT-US91-00899-12	Sequence 12, Appl
31	61.5	18.9	399	4	US-09-949-016-11353	Sequence 11353, A
32	59.5	18.3	340	2	US-08-578-592-5	Sequence 5, Appli
33	59.5	18.3	340	3	US-09-185-111-5	Sequence 5, Appli
34	58.5	17.9	365	3	US-09-151-592-2	Sequence 2, Appli
35	58.5	17.9	365	3	US-09-254-077A-10	Sequence 10, Appl
36	58.5	17.9	365	4	US-09-443-766-13	Sequence 13, Appl
37	56	17.2	7257	3	US-09-335-409-5	Sequence 5, Appli
38	56	17.2	7257	3	US-09-568-102-5	Sequence 5, Appli
39	56	17.2	7257	3	US-09-567-969-5	Sequence 5, Appli
40	56	17.2	7257	3	US-09-568-480-5	Sequence 5, Appli
41	56	17.2	7257	3	US-09-568-486-5	Sequence 5, Appli
42	56	17.2	7257	3	US-09-568-472-5	Sequence 5, Appli
43	56	17.2	7257	3	US-09-567-899-5	Sequence 5, Appli
44	55.5	17.0	340	3	US-09-254-077A-6	Sequence 6, Appli
45	55.5	17.0	340	3	US-09-254-077A-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-978-741-3
; Sequence 3, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-978-741-3

Query Match 98.2%; Score 320; DB 3; Length 61;
Best Local Similarity 100.0%; Pred. No. 1e-35;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 RLAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RLAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
```

```
                |
Db          61 H 61


RESULT 2
US-09-333-729A-5
; Sequence 5, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 5
;   LENGTH: 61
;   TYPE: PRT
;   ORGANISM: Cricetulus Grieseus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 15, 17, 38
;   OTHER INFORMATION: unknown amino acid
US-09-333-729A-5

  Query Match             98.2%;  Score 320;  DB 3;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 1e-35;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RLAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RLAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 3
US-08-978-741-9
; Sequence 9, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  61 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-9

  Query Match             90.5%;  Score 295;  DB 3;  Length 61;
  Best Local Similarity   91.5%;  Pred. No. 2.3e-32;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||
Db          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 4
US-09-333-729A-4
; Sequence 4, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 4
;   LENGTH: 61
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-333-729A-4

  Query Match             90.5%;  Score 295;  DB 3;  Length 61;
  Best Local Similarity   91.5%;  Pred. No. 2.3e-32;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||
Db          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 5
US-08-978-741-2
; Sequence 2, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
```

```
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  365 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-2

  Query Match             90.5%;  Score 295;  DB 3;  Length 365;
  Best Local Similarity   91.5%;  Pred. No. 1.8e-31;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | |||||||||||||||||||  ||||||||||||||||||||||||
Db          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 6
US-09-333-729A-3
; Sequence 3, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 3
;   LENGTH: 365
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-333-729A-3

  Query Match             90.5%;  Score 295;  DB 3;  Length 365;
  Best Local Similarity   91.5%;  Pred. No. 1.8e-31;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | |||||||||||||||||||  ||||||||||||||||||||||||
Db          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 7
US-08-978-741-6
; Sequence 6, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  397 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-6

  Query Match             90.5%;  Score 295;  DB 3;  Length 397;
  Best Local Similarity   91.5%;  Pred. No. 2e-31;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | |||||||||||||||||||  ||||||||||||||||||||||||
Db         35 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 93


RESULT 8
US-09-333-729A-7
; Sequence 7, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 7
;   LENGTH: 397
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid insert encoded protein.
US-09-333-729A-7

  Query Match             90.5%;  Score 295;  DB 3;  Length 397;
  Best Local Similarity   91.5%;  Pred. No. 2e-31;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | |||||||||||||||||||  ||||||||||||||||||||||||
Db         35 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 93


RESULT 9
US-08-978-741-17
; Sequence 17, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
```

```
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  343 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-17

  Query Match             62.6%;  Score 204;  DB 3;  Length 343;
  Best Local Similarity   94.9%;  Pred. No. 2.8e-19;
  Matches   37;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         23 NQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              |||||||||||||||  ||||||||||||||||||||||
Db          1 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 39


RESULT 10
US-09-333-729A-13
; Sequence 13, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 13
;   LENGTH: 343
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-333-729A-13

  Query Match             62.6%;  Score 204;  DB 3;  Length 343;
  Best Local Similarity   94.9%;  Pred. No. 2.8e-19;
  Matches   37;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         23 NQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              |||||||||||||||  ||||||||||||||||||||||
Db          1 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 39


RESULT 11
US-09-270-767-45921
; Sequence 45921, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45921
;   LENGTH: 399
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-270-767-45921

  Query Match             54.3%;  Score 177;  DB 4;  Length 399;
  Best Local Similarity   73.3%;  Pred. No. 1.4e-15;
  Matches   33;  Conservative    3;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          7 DLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQ 51
              |  ||| | | |||||||||||||||||||   ||| :|||:||:
Db         25 DPNGYLTYCPCMGRFGNQADHFLGSLAFAKALNRTLILPPWVEYR 69


RESULT 12
US-08-978-741-8
; Sequence 8, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  474 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-8

  Query Match             35.0%;  Score 114;  DB 3;  Length 474;
  Best Local Similarity   64.7%;  Pred. No. 4.8e-07;
  Matches   22;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         19 GRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQH 52
              |||||| | ||| |||||   ||| :| :||::|
Db        166 GRFGNQVDQFLGVLAFAKALDRTLVLPNFIEFKH 199


RESULT 13
US-09-333-729A-12
```

```
; Sequence 12, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 12
;   LENGTH: 474
;   TYPE: PRT
;   ORGANISM: Caenorhabditis Elegans
US-09-333-729A-12

  Query Match             35.0%;  Score 114;  DB 3;  Length 474;
  Best Local Similarity   64.7%;  Pred. No. 4.8e-07;
  Matches   22;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         19 GRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQH 52
              |||||| | ||| |||||   ||| :| :||::|
Db        166 GRFGNQVDQFLGVLAFAKALDRTLVLPNFIEFKH 199


RESULT 14
US-08-978-741-14
; Sequence 14, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-14

  Query Match             24.8%;  Score 81;  DB 3;  Length 28;
  Best Local Similarity   94.4%;  Pred. No. 0.00048;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGR 20
              ||||| ||||||||||||
Db         11 AGSWDPAGYLLYXPXMGR 28


RESULT 15
US-09-333-729A-14
; Sequence 14, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 14
;   LENGTH: 28
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid insert encoded protein.
;   NAME/KEY: unsure
;   LOCATION: 23, 25
;   OTHER INFORMATION: unknown amino acid
US-09-333-729A-14

  Query Match             24.8%;  Score 81;  DB 3;  Length 28;
  Best Local Similarity   94.4%;  Pred. No. 0.00048;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGR 20
              ||||| ||||||||||||
Db         11 AGSWDPAGYLLYXPXMGR 28


Search completed: October 25, 2005, 15:42:48
Job time : 11.0205 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:22:39 ; Search time 35.8234 Seconds
(without alignments)
710.936 Million cell updates/sec

Title: US-09-774-954-3
Perfect score: 326
Sequence: 1 RLAGSWDLAGYLLYXPXMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1862994 seqs, 417510619 residues

Total number of hits satisfying chosen parameters: 1862994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	320	98.2	61	11	US-09-774-954-3	Sequence 3, Appli
2	295	90.5	61	11	US-09-774-954-9	Sequence 9, Appli
3	295	90.5	365	11	US-09-774-954-2	Sequence 2, Appli
4	295	90.5	388	14	US-10-301-822-169	Sequence 169, App
5	295	90.5	397	11	US-09-774-954-6	Sequence 6, Appli
6	204	62.6	343	11	US-09-774-954-17	Sequence 17, Appl
7	177	54.3	402	20	US-11-097-143-18162	Sequence 18162, A
8	114	35.0	474	11	US-09-774-954-8	Sequence 8, Appli
9	81	24.8	28	11	US-09-774-954-14	Sequence 14, Appl
10	76	23.3	150	18	US-10-820-474A-102	Sequence 102, App
11	72.5	22.2	601	14	US-10-156-761-12551	Sequence 12551, A
12	71.5	21.9	189	16	US-10-425-115-201329	Sequence 201329,
13	69.5	21.3	198	15	US-10-282-122A-61790	Sequence 61790, A
14	68.5	21.0	152	15	US-10-424-599-266118	Sequence 266118,
15	67.5	20.7	538	16	US-10-437-963-113919	Sequence 113919,
16	66.5	20.4	741	17	US-10-472-928-1698	Sequence 1698, Ap
17	66.5	20.4	763	18	US-10-617-320-3610	Sequence 3610, Ap
18	66.5	20.4	767	9	US-09-815-242-13343	Sequence 13343, A
19	66.5	20.4	767	9	US-09-815-242-13598	Sequence 13598, A
20	66.5	20.4	767	10	US-09-769-787-96	Sequence 96, Appl
21	66.5	20.4	767	15	US-10-282-122A-73911	Sequence 73911, A
22	66.5	20.4	783	9	US-09-775-978-2	Sequence 2, Appli
23	61.5	18.9	365	9	US-09-254-077A-11	Sequence 11, Appl
24	61.5	18.9	365	9	US-09-863-475A-6	Sequence 6, Appli
25	61.5	18.9	365	14	US-10-105-963-10	Sequence 10, Appl
26	61	18.7	75	16	US-10-437-963-198684	Sequence 198684,
27	61	18.7	117	16	US-10-425-115-235644	Sequence 235644,
28	60.5	18.6	580	15	US-10-369-493-8280	Sequence 8280, Ap
29	59.5	18.3	198	15	US-10-282-122A-62362	Sequence 62362, A
30	59.5	18.3	198	15	US-10-282-122A-64747	Sequence 64747, A
31	59.5	18.3	241	15	US-10-424-599-273532	Sequence 273532,
32	59	18.1	126	15	US-10-424-599-219171	Sequence 219171,
33	59	18.1	335	14	US-10-156-761-11617	Sequence 11617, A
34	58.5	17.9	223	15	US-10-282-122A-63705	Sequence 63705, A
35	58.5	17.9	365	9	US-09-051-034A-4	Sequence 4, Appli
36	58.5	17.9	365	9	US-09-254-077A-10	Sequence 10, Appl
37	58.5	17.9	365	9	US-09-844-268-13	Sequence 13, Appl
38	58.5	17.9	365	9	US-09-844-705-13	Sequence 13, Appl
39	58	17.8	110	16	US-10-425-115-327031	Sequence 327031,
40	57.5	17.6	587	15	US-10-424-599-179849	Sequence 179849,
41	57	17.5	322	16	US-10-664-421-10	Sequence 10, Appl
42	57	17.5	322	16	US-10-377-268-18	Sequence 18, Appl
43	57	17.5	322	18	US-10-941-635-10	Sequence 10, Appl
44	57	17.5	363	15	US-10-369-493-5662	Sequence 5662, Ap
45	57	17.5	376	17	US-10-732-923-19411	Sequence 19411, A

ALIGNMENTS

RESULT 1
US-09-774-954-3
; Sequence 3, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:

```
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 3:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 61 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-774-954-3

  Query Match             98.2%;  Score 320;  DB 11;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-35;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 RLAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RLAGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 2
US-09-774-954-9
; Sequence 9, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 9:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 61 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-774-954-9

  Query Match             90.5%;  Score 295;  DB 11;  Length 61;
  Best Local Similarity   91.5%;  Pred. No. 9.2e-32;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||||
Db          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 3
US-09-774-954-2
; Sequence 2, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 2:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 365 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-774-954-2

  Query Match             90.5%;  Score 295;  DB 11;  Length 365;
  Best Local Similarity   91.5%;  Pred. No. 6.3e-31;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||||
Db          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 4
US-10-301-822-169
; Sequence 169, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
;  APPLICANT:  Burgart, Lawrence J.
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
```

```
;  CURRENT APPLICATION NUMBER: US/10/301,822
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 169
;   LENGTH: 388
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-10-301-822-169

  Query Match             90.5%;  Score 295;  DB 14;  Length 388;
  Best Local Similarity   91.5%;  Pred. No. 6.7e-31;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||
Db         26 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 84


RESULT 5
US-09-774-954-6
; Sequence 6, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 6:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 397 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-774-954-6

  Query Match             90.5%;  Score 295;  DB 11;  Length 397;
  Best Local Similarity   91.5%;  Pred. No. 6.9e-31;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||
Db         35 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 93


RESULT 6
US-09-774-954-17
; Sequence 17, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 17:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 343 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-774-954-17

  Query Match             62.6%;  Score 204;  DB 11;  Length 343;
  Best Local Similarity   94.9%;  Pred. No. 9.5e-19;
  Matches   37;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         23 NQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
              |||||||||||||||  ||||||||||||||||||||||
Db          1 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 39


RESULT 7
US-11-097-143-18162
; Sequence 18162, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
```

```
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18162
;   LENGTH: 402
;   TYPE: PRT
;   ORGANISM: DROSOPHILA
US-11-097-143-18162

  Query Match             54.3%;  Score 177;  DB 20;  Length 402;
  Best Local Similarity   73.3%;  Pred. No. 4.8e-15;
  Matches   33;  Conservative    3;  Mismatches    9;  Indels    0;  Gaps    0;

Qy          7 DLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQ 51
              |  ||| | | ||||||||||||||||||||    ||| :|||:||:
Db         28 DPNGYLTYCPCMGRFGNQADHFLGSLAFAKALNRTLILPPWVEYR 72


RESULT 8
US-09-774-954-8
; Sequence 8, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 8:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 474 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-774-954-8

  Query Match             35.0%;  Score 114;  DB 11;  Length 474;
  Best Local Similarity   64.7%;  Pred. No. 1.6e-06;
  Matches   22;  Conservative    4;  Mismatches    8;  Indels    0;  Gaps    0;

Qy         19 GRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQH 52
              |||||| | ||| |||||   ||| :| :||::|
Db        166 GRFGNQVDQFLGVLAFAKALDRTLVLPNFIEFKH 199


RESULT 9
US-09-774-954-14
; Sequence 14, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 14:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 28 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-774-954-14

  Query Match             24.8%;  Score 81;  DB 11;  Length 28;
  Best Local Similarity   94.4%;  Pred. No. 0.0021;
  Matches   17;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGR 20
              ||||| ||||||||||||
Db         11 AGSWDPAGYLLYXPXMGR 28


RESULT 10
US-10-820-474A-102
; Sequence 102, Application US/10820474A
; Publication No. US20050155089A1
; GENERAL INFORMATION:
;  APPLICANT: LAL, PREETI
;  APPLICANT:  TANG, Y. TOM
;  APPLICANT:  GORGONE, GINA A.
;  APPLICANT:  CORLEY, NEIL C.
```

```
;  APPLICANT:  GUEGLER, KARL J.
;  APPLICANT:  BAUGHN, MARIAH R.
;  APPLICANT:  AKERBLOM, INGRID E.
;  APPLICANT:  AU-YOUNG, JANICE
;  APPLICANT:  YUE, HENRY
;  APPLICANT:  PATTERSON, CHANDRA
;  APPLICANT:  REDDY, ROOPA
;  APPLICANT:  HILLMAN, JENNIFER L.
;  APPLICANT:  BANDMAN, OLGA
;  TITLE OF INVENTION: SIGNAL PEPTIDE-CONTAINING MOLECULES
;  FILE REFERENCE: 039386-1568
;  CURRENT APPLICATION NUMBER: US/10/820,474A
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: 09/720,533
;  PRIOR FILING DATE: 2001-03-20
;  PRIOR APPLICATION NUMBER: PCT/US99/14484
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/090,762
;  PRIOR FILING DATE: 1998-06-26
;  PRIOR APPLICATION NUMBER: 60/094,983
;  PRIOR FILING DATE: 1998-07-31
;  PRIOR APPLICATION NUMBER: 60/102,686
;  PRIOR FILING DATE: 1998-10-01
;  NUMBER OF SEQ ID NOS: 269
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 102
;   LENGTH: 150
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte Clone No: 1647884
US-10-820-474A-102

  Query Match             23.3%;  Score 76;  DB 18;  Length 150;
  Best Local Similarity   77.8%;  Pred. No. 0.059;
  Matches   14;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          3 AGSWDLAGYLLYXPXMGR 20
              ||||| |||||| | ||:
Db         26 AGSWDPAGYLLYCPCMGK 43


RESULT 11
US-10-156-761-12551
; Sequence 12551, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12551
;   LENGTH: 601
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-12551

  Query Match             22.2%;  Score 72.5;  DB 14;  Length 601;
  Best Local Similarity   29.9%;  Pred. No. 0.77;
  Matches   23;  Conservative    8;  Mismatches   17;  Indels   29;  Gaps    5;

Qy          5 SWDLAGYLLYXPXMGRFGNQADHFL-------------GSLAFAKLXVRTLAVPP----- 46
              |||:  | |: | :| :| |  | |              ||| :|     || |||
Db         60 SWDIKNY-LWAPRLGMYGIQRIHLLGNVMVLAGAGVGGGSLNYA----NTLYVPPKAFFD 114

Qy         47 ---W---IEYQHHKPPF 57
                 |    ::|     |:
Db        115 DPQWKDITDWQEELKPY 131


RESULT 12
US-10-425-115-201329
; Sequence 201329, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 201329
;   LENGTH: 189
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_115198C.1.pep
US-10-425-115-201329

  Query Match             21.9%;  Score 71.5;  DB 16;  Length 189;
  Best Local Similarity   40.0%;  Pred. No. 0.3;
  Matches   22;  Conservative    4;  Mismatches   24;  Indels    5;  Gaps    3;

Qy          1 RLAGSWDLAGY--LLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHH 53
              |  || :| |   | | |  | | ||      :   | |  ||| ||| ::  ||
Db         67 RCGGSVELEGRGGLWYAPHSG-FSNQVGELRNAAVAAALLNRTLVVPPVLD--HH 118


RESULT 13
US-10-282-122A-61790
; Sequence 61790, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
```

```
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61790
;   LENGTH: 198
;   TYPE: PRT
;   ORGANISM: Mycobacterium avium
US-10-282-122A-61790

  Query Match             21.3%;  Score 69.5;  DB 15;  Length 198;
  Best Local Similarity   50.0%;  Pred. No. 0.59;
  Matches   16;  Conservative    2;  Mismatches   11;  Indels    3;  Gaps    1;

Qy          21 FGNQADHFLGSLAFAKLXVRTLAV---PPWIE 49
               || | | | | :||| | |   ||    ||:|
Db         116 FGRQVDSFEGDIAFAGLDVPVRAVFIRAPWVE 147


RESULT 14
US-10-424-599-266118
; Sequence 266118, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 266118
;   LENGTH: 152
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(152)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_82325C.1.pep
US-10-424-599-266118

  Query Match             21.0%;  Score 68.5;  DB 15;  Length 152;
  Best Local Similarity   37.2%;  Pred. No. 0.61;
  Matches   16;  Conservative    7;  Mismatches   17;  Indels    3;  Gaps    2;

Qy          11 YLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHH 53
               :: | |  | | ||   |  ::  | :  ||| ||| ::  ||
Db          76 FVWYAPHSG-FSNQLSEFKNAVLMAGILNRTLVVPPILD--HH 115


RESULT 15
US-10-437-963-113919
; Sequence 113919, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 113919
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_17661C.1.pep
US-10-437-963-113919

  Query Match             20.7%;  Score 67.5;  DB 16;  Length 538;
  Best Local Similarity   38.2%;  Pred. No. 3.2;
  Matches   21;  Conservative    5;  Mismatches   24;  Indels    5;  Gaps    3;

Qy           1 RLAGSWDLAG--YLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHH 53
               |  | :|  |  :| | |  | | ||      :   | |  ||| ||| ::  ||
Db          98 RCGRSPELEGERFLWYAPHSG-FSNQVGELRNAAVAAALLNRTLVVPPVLD--HH 149


Search completed: October 25, 2005, 15:41:23
Job time : 37.8234 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:21:23 ; Search time 35.1971 Seconds
(without alignments)
887.482 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	353	100.0	388	1	OFU1_HUMAN	Q9h488 homo sapien
2	353	100.0	388	1	OFU1_PANTR	Q6ev69 pan troglod
3	343	97.2	351	2	Q7YRE6	Q7yre6 bos taurus
4	343	97.2	391	2	Q7YRE7	Q7yre7 bos taurus
5	341	96.6	378	2	Q7YRZ3	Q7yrz3 sus scrofa
6	331	93.8	393	1	OFU1_MOUSE	Q91zw2 mus musculu
7	331	93.8	393	2	Q8C8R4	Q8c8r4 mus musculu
8	327	92.6	395	2	Q6EV70	Q6ev70 rattus norv
9	326	92.4	391	2	Q6EV66	Q6ev66 bos taurus
10	308	87.3	380	2	Q8AWB4	Q8awb4 gallus gall
11	295	83.6	61	1	OFU1_CRIGR	P83337 cricetulus
12	287	81.3	395	2	Q7T028	Q7t028 brachydanio
13	284	80.5	380	2	Q640S0	Q640s0 xenopus tro
14	284	80.5	384	2	Q70AG7	Q70ag7 fugu rubrip
15	284	80.5	387	2	Q6EV65	Q6ev65 tetraodon n
16	284	80.5	396	2	Q6EV71	Q6ev71 xenopus tro
17	268	75.9	380	2	Q8AXS8	Q8axs8 xenopus lae
18	223	63.2	402	1	OFU1_DROME	Q9v6x7 drosophila
19	223	63.2	402	2	Q86SA7	Q86sa7 drosophila
20	223	63.2	402	2	Q659S1	Q659s1 drosophila
21	209	59.2	399	2	Q7QHS7	Q7qhs7 anopheles g
22	207	58.6	381	2	Q6EV67	Q6ev67 ciona savig
23	192	54.4	385	2	Q659R9	Q659r9 onchocerca
24	186	52.7	402	2	Q6EV68	Q6ev68 drosophila
25	183	51.8	381	1	OFU1_CAEEL	Q18014 caenorhabdi
26	177	50.1	434	2	Q659S2	Q659s2 caenorhabdi
27	164	46.5	364	2	Q8MSR1	Q8msr1 drosophila
28	161.5	45.8	353	2	Q659S0	Q659s0 bombyx mori
29	79.5	22.5	832	2	Q7YTA5	Q7yta5 leishmania
30	78	22.1	354	1	OPSI_ASTFA	P51474 astyanax fa
31	77.5	22.0	538	2	Q7XC59	Q7xc59 oryza sativ
32	77.5	22.0	538	2	Q9AUT5	Q9aut5 oryza sativ
33	76.5	21.7	602	2	Q9L0H6	Q910h6 streptomyce
34	73.5	20.8	601	2	Q82DG5	Q82dg5 streptomyce
35	73	20.7	355	2	Q9IA35	Q9ia35 poephila gu
36	73	20.7	507	2	Q9HGR6	Q9hgr6 mucor circi
37	71	20.1	310	2	Q9I973	Q9i973 serinus can
38	71	20.1	349	1	OPSG_BRARE	Q9w6a5 brachydanio
39	71	20.1	349	2	Q8AYM9	Q8aym9 brachydanio
40	70.5	20.0	227	2	Q90WD5	Q90wd5 hippoglossu
41	70.5	20.0	339	2	Q8QGR3	Q8qgr3 hippoglossu
42	70.5	20.0	349	2	Q801U8	Q801u8 brachydanio
43	70.5	20.0	349	2	Q8AYM8	Q8aym8 brachydanio
44	70.5	20.0	902	2	Q962E4	Q962e4 leishmania
45	69.5	19.7	1532	2	Q9P2K1	Q9p2k1 homo sapien

ALIGNMENTS

RESULT 1
OFU1_HUMAN
ID OFU1_HUMAN STANDARD; PRT; 388 AA.
AC Q9H488; Q14685; Q9BW76;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)
DE (Peptide O-fucosyltransferase) (O-FucT-1).
GN Name=POFUT1; Synonyms=KIAA0180;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, AND INDUCTION.
RC TISSUE=Heart;
RX MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;
RA Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,
RA Haltiwanger R.S.;
RT "Modification of epidermal growth factor-like repeats with O-fucose:
RT molecular cloning and expression of a novel GDP-fucose protein O-
RT fucosyltransferase.";
RL J. Biol. Chem. 276:40338-40345(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual
RT curation of 330 KIAA cDNA clones.";
RL DNA Res. 9:99-106(2002).
RN [3]
RP SEQUENCE OF 46-388 FROM N.A. (ISOFORM 1).
RC TISSUE=Bone marrow;
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 3:17-24(1996).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052; DOI=10.1038/414865a;
RA Deloukas P., Matthews L.H., Ashurst J.L., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,

check the date

```
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaeslaiho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [6]
RP   CHARACTERIZATION.
RX   MEDLINE=97175972; PubMed=9023546;
RA   Wang Y., Lee G.F., Kelley R.F., Spellman M.W.;
RT   "Identification of a GDP-L-fucose:polypeptide fucosyltransferase and
RT   enzymatic addition of O-linked fucose to EGF domains.";
RL   Glycobiology 6:837-842(1996).
CC   -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC       O-glycosidic linkage to a conserved serine or threonine residue in
CC       EGF domains.
CC   -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC       beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC   -!- COFACTOR: Manganese.
CC   -!- PATHWAY: Glycosylation.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9H488-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9H488-2; Sequence=VSP_001809;
CC         Note=No experimental confirmation available;
CC   -!- TISSUE SPECIFICITY: Highly expressed in heart, brain, placenta,
CC       lung, liver, skeletal muscle, kidney and pancreas.
CC   -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF375884; AAL09576.1; -.
DR   EMBL; D80002; BAA11497.2; -.
DR   EMBL; AL121897; CAC16424.1; -.
DR   EMBL; BC000582; AAH00582.1; -.
DR   Genew; HGNC:14988; POFUT1.
DR   H-InvDB; HIX0015724; -.
DR   MIM; 607491; -.
DR   GO; GO:0005737; C:cytoplasm; ISS.
DR   GO; GO:0030173; C:integral to Golgi membrane; NAS.
DR   GO; GO:0046922; F:peptide-O-fucosyltransferase activity; TAS.
DR   GO; GO:0009790; P:embryonic development; NAS.
DR   GO; GO:0007219; P:Notch signaling pathway; ISS.
DR   GO; GO:0016266; P:O-glycan processing; TAS.
DR   GO; GO:0045449; P:regulation of transcription; NAS.
DR   GO; GO:0006445; P:regulation of translation; NAS.
KW   Alternative splicing; Fucose metabolism; Glycoprotein;
KW   Glycosyltransferase; Manganese; Signal; Transferase.
FT   SIGNAL        1     26       Potential.
FT   CHAIN        27    388       GDP-fucose protein O-fucosyltransferase
FT                                 1.
FT   CARBOHYD     62     62       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    160    160       N-linked (GlcNAc. . .) (Potential).
FT   VARSPLIC    182    388       FSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEMVKTG
FT                                EAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMA
FT                                SPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQ
FT                                SVYVATDSESYVPELQQLFKGKVKVVSLKPEVAQVDLYILG
FT                                QADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRD
FT                                EF -> RENHSCVTLLFPR (in isoform 2).
FT                                /FTId=VSP_001809.
SQ   SEQUENCE   388 AA;  43955 MW;  3FACCCA434D02415 CRC64;

  Query Match             100.0%;  Score 353;  DB 1;  Length 388;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-36;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 83

Qy         61 H 61
              |
Db         84 H 84


RESULT 2
OFU1_PANTR
ID   OFU1_PANTR     STANDARD;      PRT;   388 AA.
AC   Q6EV69;
DT   25-OCT-2004 (Rel. 45, Created)
DT   25-OCT-2004 (Rel. 45, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)
DE   (Peptide O-fucosyltransferase) (O-FucT-1).
GN   Name=POFUT1; Synonyms=FUT12;
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=12966037; DOI=10.1093/glycob/cwg113;
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RA   Oriol R.;
RT   "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT   fucosyltransferases and alpha6-fucosyltransferases: phylogeny and
RT   identification of conserved peptide motifs.";
RL   Glycobiology 13:1C-5C(2003).
CC   -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC       O-glycosidic linkage to a conserved serine or threonine residue in
CC       EGF domains (By similarity).
CC   -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC       beta-L-fucose to the serine hydroxy group of a protein acceptor.
```

```
CC   -!- COFACTOR: Manganese (By similarity).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ781500; CAH03712.1; -.
KW   Fucose metabolism; Glycoprotein; Glycosyltransferase; Manganese;
KW   Signal; Transferase.
FT   SIGNAL        1     26       Potential.
FT   CHAIN        27    388       GDP-fucose protein O-fucosyltransferase
FT                                 1.
FT   CARBOHYD     62     62       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    160    160       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   388 AA;  43955 MW;  3FACCCA434D02415 CRC64;

  Query Match             100.0%;  Score 353;  DB 1;  Length 388;
  Best Local Similarity   100.0%;  Pred. No. 4.5e-36;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 83

Qy         61 H 61
              |
Db         84 H 84


RESULT 3
Q7YRE6
ID   Q7YRE6      PRELIMINARY;      PRT;   351 AA.
AC   Q7YRE6;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Protein O-fucosyltransferase 1b.
GN   Name=pofut1;
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Loriol C., Germot A., Dupuy F., Maftah A.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY344581; AAQ02333.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   351 AA;  39576 MW;  BF830F61A7296F42 CRC64;

  Query Match             97.2%;  Score 343;  DB 2;  Length 351;
  Best Local Similarity   95.1%;  Pred. No. 7.5e-35;
  Matches   58;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              :| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         27 LPVGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNV 86

Qy         61 H 61
              |
Db         87 H 87


RESULT 4
Q7YRE7
ID   Q7YRE7      PRELIMINARY;      PRT;   391 AA.
AC   Q7YRE7;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Protein O-fucosyltransferase 1a.
GN   Name=pofut1;
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Loriol C., Germot A., Dupuy F., Maftah A.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY344580; AAQ02332.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   391 AA;  44307 MW;  96762AB81A2027AD CRC64;

  Query Match             97.2%;  Score 343;  DB 2;  Length 391;
  Best Local Similarity   95.1%;  Pred. No. 8.4e-35;
  Matches   58;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              :| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db         27 LPVGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNV 86

Qy         61 H 61
              |
Db         87 H 87


RESULT 5
Q7YRZ3
ID   Q7YRZ3      PRELIMINARY;      PRT;   378 AA.
AC   Q7YRZ3;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Protein-O-fucosyltransferase (Fragment).
GN   Name=POFUT1;
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Martinez-Dunker I., Mollicone R., Candelier J.J., Breton C., Oriol R.;
RT   "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT   fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and
RT   identification of conserved peptide motifs.";
RL   Glycobiology 13:1c-5c(2003).
DR   EMBL; AJ567917; CAD99201.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
FT   NON_TER       1      1
FT   NON_TER     378    378
SQ   SEQUENCE   378 AA;  42842 MW;  1870DEB0C50713F1 CRC64;

  Query Match             96.6%;  Score 341;  DB 2;  Length 378;
  Best Local Similarity   96.7%;  Pred. No. 1.5e-34;
  Matches   59;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              :|| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 LPAVSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 79

Qy         61 H 61
              |
Db         80 H 80
```

```
RESULT 6
OFU1_MOUSE
ID   OFU1_MOUSE     STANDARD;      PRT;   393 AA.
AC   Q91ZW2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   GDP-fucose protein O-fucosyltransferase 1 precursor (EC 2.4.1.221)
DE   (Peptide O-fucosyltransferase) (O-FucT-1).
GN   Name=Pofut1;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SvJ; TISSUE=Liver;
RX   MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;
RA   Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,
RA   Haltiwanger R.S.;
RT   "Modification of epidermal growth factor-like repeats with O-fucose:
RT   molecular cloning and expression of a novel GDP-fucose protein O-
RT   fucosyltransferase.";
RL   J. Biol. Chem. 276:40338-40345(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N-3; TISSUE=Mammary gland;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC       O-glycosidic linkage to a conserved serine or threonine residue in
CC       EGF domains (By similarity).
CC   -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC       beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC   -!- COFACTOR: Manganese (By similarity).
CC   -!- PATHWAY: Glycosylation.
CC   -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF375885; AAL09577.1; -.
DR   EMBL; BC046295; AAH46295.1; -.
DR   MGD; MGI:2153207; Pofut1.
DR   GO; GO:0030173; C:integral to Golgi membrane; ISS.
DR   GO; GO:0008417; F:fucosyltransferase activity; ISS.
DR   GO; GO:0001525; P:angiogenesis; IMP.
DR   GO; GO:0007507; P:heart development; IMP.
DR   GO; GO:0007399; P:neurogenesis; IMP.
DR   GO; GO:0007219; P:Notch signaling pathway; IMP.
DR   GO; GO:0006493; P:O-linked glycosylation; ISS.
DR   GO; GO:0001756; P:somitogenesis; IMP.
KW   Fucose metabolism; Glycoprotein; Glycosyltransferase; Manganese;
KW   Signal; Transferase.
FT   SIGNAL        1     30       Potential.
FT   CHAIN        31    393       GDP-fucose protein O-fucosyltransferase
FT                                 1.
FT   CARBOHYD     67     67       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    165    165       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   393 AA;  44688 MW;  D982104E95E5CF3B CRC64;

  Query Match             93.8%;  Score 331;  DB 1;  Length 393;
  Best Local Similarity   98.3%;  Pred. No. 2.8e-33;
  Matches   58;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 AGSWDLAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 89


RESULT 7
Q8C8R4
ID   Q8C8R4      PRELIMINARY;      PRT;   393 AA.
AC   Q8C8R4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Mus musculus adult retina cDNA, RIKEN full-length enriched library,
DE   clone:A930028F21 product:protein O-fucosyltransferase 1, full insert
DE   sequence.
GN   Name=Pofut1;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
```

```
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Retina;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA   Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK044629; BAC32009.1; -.
DR   MGD; MGI:2153207; Pofut1.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0001525; P:angiogenesis; IMP.
DR   GO; GO:0006004; P:fucose metabolism; TAS.
DR   GO; GO:0007507; P:heart development; IMP.
DR   GO; GO:0007399; P:neurogenesis; IMP.
DR   GO; GO:0007219; P:Notch signaling pathway; IMP.
DR   GO; GO:0001756; P:somitogenesis; IMP.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   393 AA;  44704 MW;  1CA94CB30B85D6DC CRC64;

  Query Match             93.8%;  Score 331;  DB 2;  Length 393;
  Best Local Similarity   98.3%;  Pred. No. 2.8e-33;
  Matches   58;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         31 AGSWDLAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 89


RESULT 8
Q6EV70
ID   Q6EV70      PRELIMINARY;      PRT;   395 AA.
AC   Q6EV70;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Protein-O-fucosyltransferase 1.
GN   Name=fut12;
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar;
RX   PubMed=12966037; DOI=10.1093/glycob/cwg113;
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RA   Oriol R.;
RT   "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT   fucosyltransferases and alpha6-fucosyltransferases: phylogeny and
RT   identification of conserved peptide motifs.";
RL   Glycobiology 13:1C-5C(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Wistar;
RA   Martinez-Duncker I., Oriol R., Mollicone R.;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ781499; CAH03711.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   395 AA;  44983 MW;  2A85521E22F38920 CRC64;

  Query Match             92.6%;  Score 327;  DB 2;  Length 395;
  Best Local Similarity   98.3%;  Pred. No. 9e-33;
  Matches   57;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          4 GSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         34 GSWDLAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 91


RESULT 9
Q6EV66
ID   Q6EV66      PRELIMINARY;      PRT;   391 AA.
AC   Q6EV66;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Protein-O-fucosyltransferase 1.
GN   Name=fut12;
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=12966037; DOI=10.1093/glycob/cwg113;
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RA   Oriol R.;
RT   "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT   fucosyltransferases and alpha6-fucosyltransferases: phylogeny and
RT   identification of conserved peptide motifs.";
RL   Glycobiology 13:1C-5C(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Martinez-Duncker I., Oriol R., Mollicone R.;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ781503; CAH03714.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   391 AA;  44277 MW;  4E047D6BB5CDF7FA CRC64;

  Query Match             92.4%;  Score 326;  DB 2;  Length 391;
  Best Local Similarity   88.5%;  Pred. No. 1.2e-32;
  Matches   54;  Conservative    5;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              :| |||:||||||:||||||||||||||||||| |:||||||||||||||||||:
Db         27 LPVGSWNPAGYLLFCPCMGRFGNQADHFLGSLAFAPLMNRTLAVPPWIEYQHHKPPFTNV 86

Qy         61 H 61
              |
Db         87 H 87


RESULT 10
Q8AWB4
ID   Q8AWB4      PRELIMINARY;      PRT;   380 AA.
AC   Q8AWB4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein POFUT1.
GN   Name=POFUT1;
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Oriol R.;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ535754; CAD59738.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   380 AA;  43316 MW;  83B53E342511811B CRC64;

  Query Match             87.3%;  Score 308;  DB 2;  Length 380;
  Best Local Similarity   86.4%;  Pred. No. 2.2e-30;
  Matches   51;  Conservative    6;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||||||||||||:||||:||||: |||||||||||||:||:||:||||
Db         18 AGSWDTAGYLLYCPCMGRFGNQAEHFLGALAFARALNRTLAVPPWIEYRHHRPPYTNLH 76


RESULT 11
OFU1_CRIGR
ID   OFU1_CRIGR     STANDARD;      PRT;    61 AA.
AC   P83337;
DT   25-OCT-2004 (Rel. 45, Created)
DT   25-OCT-2004 (Rel. 45, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   GDP-fucose protein O-fucosyltransferase 1 (EC 2.4.1.221) (Peptide O-
DE   fucosyltransferase) (O-FucT-1) (Fragment).
GN   Name=POFUT1;
OS   Cricetulus griseus (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetulus.
OX   NCBI_TaxID=10029;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=21523965; PubMed=11524432; DOI=10.1074/jbc.M107849200;
RA   Wang Y., Shao L., Shi S., Harris R.J., Spellman M.W., Stanley P.,
RA   Haltiwanger R.S.;
RT   "Modification of epidermal growth factor-like repeats with O-fucose:
RT   molecular cloning and expression of a novel GDP-fucose protein O-
RT   fucosyltransferase.";
RL   J. Biol. Chem. 276:40338-40345(2001).
RN   [2]
RP   FUNCTION.
RX   MEDLINE=93364082; PubMed=8358148;
RA   Harris R.J., Spellman M.W.;
RT   "O-linked fucose and other post-translational modifications unique to
RT   EGF modules.";
RL   Glycobiology 3:219-224(1993).
RN   [3]
RP   FUNCTION.
RX   MEDLINE=97175972; PubMed=9023546;
RA   Wang Y., Lee G.F., Kelley R.F., Spellman M.W.;
RT   "Identification of a GDP-L-fucose:polypeptide fucosyltransferase and
RT   enzymatic addition of O-linked fucose to EGF domains.";
RL   Glycobiology 6:837-842(1996).
RN   [4]
RP   FUNCTION, AND GLYCOSYLATION.
RC   TISSUE=Ovary;
RX   MEDLINE=98192601; PubMed=9525914; DOI=10.1074/jbc.273.14.8112;
RA   Wang Y., Spellman M.W.;
RT   "Purification and characterization of a GDP-fucose:polypeptide
RT   fucosyltransferase from Chinese hamster ovary cells.";
RL   J. Biol. Chem. 273:8112-8118(1998).
CC   -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC       O-glycosidic linkage to a conserved serine or threonine residue in
CC       EGF domains.
CC   -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC       beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC   -!- COFACTOR: Manganese. Other divalent cations increase activity:
CC       calcium, cobalt, cadmium, magnesium and nickel.
CC   -!- PATHWAY: Glycosylation.
CC   -!- PTM: N-glycosylated.
CC   -!- SIMILARITY: Belongs to the glycosyltransferase 68 family.
KW   Direct protein sequencing; Fucose metabolism; Glycoprotein;
KW   Glycosyltransferase; Manganese; Transferase.
FT   NON_TER      61      61
SQ   SEQUENCE   61 AA;  6951 MW;  E9507AE60018F23A CRC64;

  Query Match             83.6%;  Score 295;  DB 1;  Length 61;
  Best Local Similarity   91.5%;  Pred. No. 1.5e-29;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||
Db          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 12
Q7T028
ID   Q7T028      PRELIMINARY;      PRT;   395 AA.
AC   Q7T028;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   GDP-fucose protein O-fucosyltransferase 1 (EC 2.4.1.221).
GN   Name=pofut1; Synonyms=fut12;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RA   Oriol R.;
RL   Submitted (JUL-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ579536; CAE18459.1; -.
DR   ZFIN; ZDB-GENE-040303-2; pofut1.
DR   GO; GO:0046922; F:peptide-O-fucosyltransferase activity; IEA.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   395 AA;  45165 MW;  15BBE06172542E8A CRC64;

  Query Match             81.3%;  Score 287;  DB 2;  Length 395;
  Best Local Similarity   82.5%;  Pred. No. 1e-27;
  Matches   47;  Conservative    5;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          5 SWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              :||  ||:||||||||||| ||||||||||:|||||||||| |:|| |||||:|
Db         33 TWDENGYILYCPCMGRFGNQVDHFLGSLAFAKMLNRTLAVPPWIVYRHHSPPFTNVH 89


RESULT 13
Q640S0
ID   Q640S0      PRELIMINARY;      PRT;   380 AA.
AC   Q640S0;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein.
OS   Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8364;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
```

```
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Klein S., Gerhard D.S.;
RL   Submitted (SEP-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC082519; AAH82519.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   380 AA;  43570 MW;  8BB938CE776D5238 CRC64;

  Query Match             80.5%;  Score 284;  DB 2;  Length 380;
  Best Local Similarity   80.4%;  Pred. No. 2.4e-27;
  Matches   45;  Conservative    7;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          6 WDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              || :||:|||||||||||||||||||||||::|||| ||||| | ||:||:||:|
Db         21 WDSSGYILYCPCMGRFGNQADHFLGSLAFAKMVNRTLVVPPWIVYNHHRPPYTNVH 76


RESULT 14
Q70AG7
ID   Q70AG7      PRELIMINARY;      PRT;   384 AA.
AC   Q70AG7;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   Protein-O-fucosyltransferase.
GN   Name=fut12;
OS   Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetradontoidea; Tetraodontidae; Takifugu.
OX   NCBI_TaxID=31033;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Oriol R.;
RL   Submitted (NOV-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ606070; CAE54305.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR   InterPro; IPR000886; ER_target_S.
DR   PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   384 AA;  43650 MW;  7CAA6ADC90F40032 CRC64;

  Query Match             80.5%;  Score 284;  DB 2;  Length 384;
  Best Local Similarity   83.9%;  Pred. No. 2.4e-27;
  Matches   47;  Conservative    5;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          6 WDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              ||  ||:|||||||||||||||||||||||:||||||||||| |:|| ||:||:|
Db         23 WDQNGYVLYCPCMGRFGNQADHFLGSLAFAKMLNRTLAVPPWIVYRHHVPPYTNVH 78


RESULT 15
Q6EV65
ID   Q6EV65      PRELIMINARY;      PRT;   387 AA.
AC   Q6EV65;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Protein-O-fucosyltransferase 1.
GN   Name=fut12;
OS   Tetraodon nigroviridis (Green puffer).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC   Tetradontoidea; Tetraodontidae; Tetraodon.
OX   NCBI_TaxID=99883;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   PubMed=12966037; DOI=10.1093/glycob/cwg113;
RA   Martinez-Duncker I., Mollicone R., Candelier J.J., Breton C.,
RA   Oriol R.;
RT   "A new superfamily of protein-O-fucosyltransferases, alpha2-
RT   fucosyltransferases and alpha6-fucosyltransferases: phylogeny and
RT   identification of conserved peptide motifs.";
RL   Glycobiology 13:1C-5C(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Martinez-Duncker I., Oriol R., Mollicone R.;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ781504; CAH03715.1; -.
DR   GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
KW   Glycosyltransferase; Transferase.
SQ   SEQUENCE   387 AA;  43713 MW;  BE33E3B78C883480 CRC64;

  Query Match             80.5%;  Score 284;  DB 2;  Length 387;
  Best Local Similarity   82.5%;  Pred. No. 2.4e-27;
  Matches   47;  Conservative    5;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          5 SWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              |||  ||:|||||||||||||||||||||||:|||||||||| |:|| ||: |:|
Db         22 SWDQNGYVLYCPCMGRFGNQADHFLGSLAFAKMLNRTLAVPPWIVYRHHAPPYANVH 78


Search completed: October 25, 2005, 15:35:25
Job time : 36.1971 secs
```

Against DNA

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 05:08:04 ; Search time 1382.21 Seconds
(without alignments)
2138.440 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105432_6389/app_query.fasta_1.917
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954_@CGN_1_1_4968_@runat_25102005_105432_6389 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	353	100.0	1100	6	AR163457	AR163457 Sequence
2	353	100.0	1100	6	BD103196	BD103196 O-fucosyl
3	353	100.0	1167	6	CQ727777	CQ727777 Sequence
4	353	100.0	1167	9	AJ781500	AJ781500 Pan trogl
5	353	100.0	1300	6	AR163452	AR163452 Sequence
6	353	100.0	1300	6	BD103189	BD103189 O-fucosyl
7	353	100.0	1514	6	AR163451	AR163451 Sequence
8	353	100.0	1514	6	BD103188	BD103188 O-fucosyl
9	353	100.0	1524	9	BC000582	BC000582 Homo sapi
10	353	100.0	4560	9	HSM803330	AL832023 Homo sapi
11	353	100.0	5189	9	D80002	D80002 Homo sapien
12	353	100.0	5249	9	AF375884	AF375884 Homo sapi
13	353	100.0	5266	6	AX780153	AX780153 Sequence
14	353	100.0	11284	6	BD103190	BD103190 O-fucosyl
15	343	97.2	1056	4	AY344581	AY344581 Bos tauru
16	343	97.2	1176	4	AY344580	AY344580 Bos tauru
17	341	96.6	1134	4	SSC567917	AJ567917 Sus scrof
18	331	93.8	1302	10	AF375885	AF375885 Mus muscu
19	331	93.8	1529	10	BC046295	BC046295 Mus muscu
20	327	92.6	1188	10	AJ781499	AJ781499 Rattus no
21	326	92.4	1176	4	AJ781503	AJ781503 Bos tauru
22	308	87.3	920	5	AJ719585	AJ719585 Gallus ga
23	308	87.3	1732	5	GGA535754	AJ535754 Gallus ga
24	308	87.3	2546	5	AJ720352	AJ720352 Gallus ga
25	287	81.3	1549	5	DRE579536	AJ579536 Danio rer
26	284	80.5	1155	5	AJ606070	AJ606070 Fugu rubr
27	284	80.5	1164	5	AJ781504	AJ781504 Tetraodon
28	284	80.5	1502	5	BC082519	BC082519 Xenopus t
29	284	80.5	1538	5	AJ781498	AJ781498 Silurana
30	268	75.9	1505	5	XLA514425	AJ514425 Xenopus l
31	233	66.0	145414	9	HSA392M18	AL121897 Human DNA
32	233	66.0	200329	10	AC078911	AC078911 Mus muscu
33	233	66.0	208614	10	AL807380	AL807380 Mouse DNA
c 34	233	66.0	240336	2	AC094844	AC094844 Rattus no
c 35	233	66.0	240510	2	AC134371	AC134371 Rattus no
36	223	63.2	1209	3	AJ831490	AJ831490 Drosophil
37	223	63.2	1209	6	CQ590403	CQ590403 Sequence
38	223	63.2	1295	3	AY118651	AY118651 Drosophil
39	223	63.2	1320	6	AR509393	AR509393 Sequence
40	223	63.2	1579	3	AB093572	AB093572 Drosophil
41	218	61.8	5009	6	AR163453	AR163453 Sequence
42	218	61.8	5009	6	BD103191	BD103191 O-fucosyl
43	207	58.6	1146	3	AJ781502	AJ781502 Ciona sav
44	200	56.7	1841	3	AK112708	AK112708 Ciona int
45	192	54.4	1160	3	AJ831492	AJ831492 Onchocerc

ALIGNMENTS

RESULT 1
AR163457
LOCUS AR163457 1100 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 16 from patent US 6270987.
ACCESSION AR163457
VERSION AR163457.1 GI:16234062
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1100)
AUTHORS Wang,Y. and Spellman,M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: US 6270987-A 16 07-AUG-2001;
FEATURES Location/Qualifiers
source 1. .1100
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Alignment Scores:

Pred. No.:	1.12e-35	Length:	1100
Score:	353.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

```
US-09-774-954-9 (1-61) x AR163457 (1-1100)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 His 61
              |||
Db        181 CAT 183

RESULT 2
BD103196
LOCUS       BD103196               1100 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103196
VERSION     BD103196.1  GI:22648770
KEYWORDS    JP 2001527389-A/9.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 9 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/9
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransfease
            FH   Key             Location/Qualifiers
            FT   source          1. .1100
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1100
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

Alignment Scores:
Pred. No.:              1.12e-35        Length:         1100
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-774-954-9 (1-61) x BD103196 (1-1100)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 His 61
              |||
Db        181 CAT 183

RESULT 3
CQ727777
LOCUS       CQ727777               1167 bp    DNA     linear   PAT 03-FEB-2004
DEFINITION  Sequence 13711 from Patent WO02068579.
ACCESSION   CQ727777
VERSION     CQ727777.1  GI:42294771
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
  TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 13711 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1. .1167
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:              1.19e-35        Length:         1167
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-774-954-9 (1-61) x CQ727777 (1-1167)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 129

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 189

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 249

Qy         61 His 61
              |||
Db        250 CAT 252

RESULT 4
AJ781500
LOCUS       AJ781500               1167 bp    mRNA    linear   PRI 07-JUL-2004
DEFINITION  Pan troglodytes mRNA for protein-O-fucosyltransferase 1 (fut12
            gene).
ACCESSION   AJ781500
VERSION     AJ781500.1  GI:50057069
KEYWORDS    fut12 gene; pofut1 gene; protein-O-fucosyltransferase 1.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Martinez-Duncker,I., Mollicone,R., Candelier,J.J., Breton,C. and
            Oriol,R.
  TITLE     A new superfamily of protein-O-fucosyltransferases,
```

alpha2-fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and identification of conserved peptide motifs
JOURNAL Glycobiology 13 (12), 1C-5C (2003)
PUBMED 12966037
REFERENCE 2
AUTHORS Martinez-Duncker,I., Oriol,R. and Mollicone,R.
TITLE Phylogeny of fucosyltransferases
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1167)
AUTHORS Oriol,R.
TITLE Direct Submission
JOURNAL Submitted (06-JUL-2004) Oriol R., U504, Inserm, 16 Av. Paul Vaillant-Couturier, 94807, FRANCE

FEATURES Location/Qualifiers
source 1. .1167
/organism="Pan troglodytes"
/mol_type="mRNA"
/db_xref="taxon:9598"
gene 1. .1167
/gene="fut12"
CDS 1. .1167
/gene="fut12"
/function="adds fucose to Ser or Thr"
/note="pofut1"
/codon_start=1
/product="protein-O-fucosyltransferase 1"
/protein_id="CAH03712.1"
/db_xref="GI:50057070"
/translation="MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRF
GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRV
ISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFN
KSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEM
VKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA
PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKP
EVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF"

ORIGIN

Alignment Scores:
Pred. No.: 1.19e-35 Length: 1167
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-774-954-9 (1-61) x AJ781500 (1-1167)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 129

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 189

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 249

Qy         61 His 61
              |||
Db        250 CAT 252
```

RESULT 5
AR163452
LOCUS AR163452 1300 bp DNA linear PAT 17-OCT-2001
DEFINITION Sequence 6 from patent US 6270987.
ACCESSION AR163452
VERSION AR163452.1 GI:16234057
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.

REFERENCE 1 (bases 1 to 1300)
AUTHORS Wang,Y. and Spellman,M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: US 6270987-A 6 07-AUG-2001;

FEATURES Location/Qualifiers
source 1. .1300
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. No.: 1.33e-35 Length: 1300
Score: 353.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-774-954-9 (1-61) x AR163452 (1-1300)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy         61 His 61
              |||
Db        316 CAT 318
```

RESULT 6
BD103189
LOCUS BD103189 1300 bp DNA linear PAT 27-AUG-2002
DEFINITION O-fucosyltransferase.
ACCESSION BD103189
VERSION BD103189.1 GI:22648763
KEYWORDS JP 2001527389-A/2.
SOURCE unidentified
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 1300)
AUTHORS Wang,Y. and Spellman,M.W.
TITLE O-fucosyltransferase
JOURNAL Patent: JP 2001527389-A 2 25-DEC-2001;
GENENTECH INC
COMMENT OS Unidentified
PN JP 2001527389-A/2
PD 25-DEC-2001
PF 17-DEC-1997 JP 1998532877
PR 31-JAN-1997 US 08/792498,26-NOV-1997 US 08/978741 PI
YANG WANG,MICHAEL W SPELLMAN
PC C12N15/54,C12N9/10,C07K16/40
CC Strandedness: Single;
CC Topology: Linear;
CC O-fucosyltransferase
FH Key Location/Qualifiers
FT source 1. .1300
FT /organism='Unidentified'.

FEATURES Location/Qualifiers
source 1. .1300
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

ORIGIN

Alignment Scores:
Pred. No.: 1.33e-35 Length: 1300

```
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-774-954-9 (1-61) x BD103189 (1-1300)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy         61 His 61
              |||
Db        316 CAT 318

RESULT 7
AR163451
LOCUS       AR163451                1514 bp    DNA     linear   PAT 17-OCT-2001
DEFINITION  Sequence 2 from patent US 6270987.
ACCESSION   AR163451
VERSION     AR163451.1  GI:16234056
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: US 6270987-A 2 07-AUG-2001;
FEATURES             Location/Qualifiers
     source          1. .1514
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

Alignment Scores:
Pred. No.:              1.57e-35        Length:         1514
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-774-954-9 (1-61) x AR163451 (1-1514)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 His 61
              |||
Db        181 CAT 183

RESULT 8
BD103188
LOCUS       BD103188                1514 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103188
VERSION     BD103188.1  GI:22648762
KEYWORDS    JP 2001527389-A/1.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 1514)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 1 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/1
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .1514
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .1514
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

Alignment Scores:
Pred. No.:              1.57e-35        Length:         1514
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     6               Gaps:           0

US-09-774-954-9 (1-61) x BD103188 (1-1514)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 His 61
              |||
Db        181 CAT 183

RESULT 9
BC000582
LOCUS       BC000582                1524 bp    mRNA    linear   PRI 29-JUN-2004
DEFINITION  Homo sapiens protein O-fucosyltransferase 1, transcript variant 2,
            mRNA (cDNA clone MGC:2482 IMAGE:3162533), complete cds.
ACCESSION   BC000582
VERSION     BC000582.2  GI:37588976
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1524)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
```

```
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1524)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-NOV-2000) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     On Oct 8, 2003 this sequence version replaced gi:12653608.
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 5 Row: o Column: 9
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 27436890.
FEATURES             Location/Qualifiers
     source          1. .1524
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:2482 IMAGE:3162533"
                     /tissue_type="Brain, neuroblastoma"
                     /clone_lib="NIH_MGC_19"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1. .1524
                     /gene="POFUT1"
                     /note="synonyms: KIAA0180, O-FUT, O-Fuc-T, FUT12, MGC2482,
                     O-FucT-1"
                     /db_xref="LocusID:23509"
                     /db_xref="MIM:607491"
     CDS             50. .634
                     /gene="POFUT1"
                     /codon_start=1
                     /product="protein O-fucosyltransferase 1, isoform 2
                     precursor"
                     /protein_id="AAH00582.1"
                     /db_xref="GI:12653609"
                     /db_xref="LocusID:23509"
                     /db_xref="MIM:607491"
                     /translation="MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRF
                     GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRV
                     ISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFN
                     KSELFTGISFSASYREQWSQRRENHSCVTLLFPR"
ORIGIN

Alignment Scores:
Pred. No.:                1.58e-35        Length:          1524
Score:                    353.00          Matches:         61
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       9               Gaps:            0

US-09-774-954-9 (1-61) x BC000582 (1-1524)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy         61 His 61
              |||
Db        299 CAT 301

RESULT 10
HSM803330
LOCUS       HSM803330               4560 bp    mRNA    linear   PRI 13-MAY-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp451J114 (from clone DKFZp451J114).
ACCESSION   AL832023
VERSION     AL832023.1  GI:21732563
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 4560)
  AUTHORS   Wambutt,R., Heubner,D., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            This clone (DKFZp451J114) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1. .4560
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp451J114"
                     /db_xref="taxon:9606"
                     /clone="DKFZp451J114"
                     /tissue_type="human skeletal muscle"
                     /clone_lib="451 (synonym: hlcc1). Vector pSport1; host
                     DH10B; sites NotI + SalI"
```

```
                     /dev_stage="adult"
     polyA_site      4538
ORIGIN

Alignment Scores:
Pred. No.:              5.09e-35        Length:         4560
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-774-954-9 (1-61) x HSM803330 (1-4560)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         51 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 110

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        111 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 170

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        171 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 230

Qy         61 His 61
              |||
Db        231 CAT 233

RESULT 11
D80002
LOCUS       D80002                  5189 bp    mRNA    linear   PRI 10-JAN-2003
DEFINITION  Homo sapiens KIAA0180 mRNA, complete cds.
ACCESSION   D80002
VERSION     D80002.2  GI:20521837
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Nagase,T., Seki,N., Ishikawa,K., Tanaka,A. and Nomura,N.
  TITLE     Prediction of the coding sequences of unidentified human genes. V.
            The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
            analysis of cDNA clones from human cell line KG-1
  JOURNAL   DNA Res. 3 (1), 17-24 (1996)
  MEDLINE   96281124
   PUBMED   8724849
REFERENCE   2
  AUTHORS   Chiang,P.W., Wang,S., Smithivas,P., Song,W.J., Ramamoorthy,S.,
            Hillman,J., Puett,S., Van Keuren,M.L., Crombez,E., Kumar,A.,
            Glover,T.W., Miller,D.E., Tsai,C.H., Blackburn,C.C., Chen,X.N.,
            Sun,Z., Cheng,J.F., Korenberg,J.R. and Kurnit,D.M.
  TITLE     Identification and analysis of the human and murine putative
            chromatin structure regulator SUPT6H and Supt6h
  JOURNAL   Genomics 34 (3), 328-333 (1996)
  MEDLINE   96374824
   PUBMED   8786132
REFERENCE   3  (bases 1 to 5189)
  AUTHORS   Ohara,O., Nagase,T., Kikuno,R. and Nomura,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-1995) Osamu Ohara, Kazusa DNA Research Institute;
            1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913)
COMMENT     On May 9, 2002 this sequence version replaced gi:1136419.
FEATURES             Location/Qualifiers
     source          1. .5189
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="ha02567s1"
                     /sex="male"
                     /cell_line="KG-1"
                     /cell_type="myeloblast"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
                     /note="This sequence was obtained by subcloning of the DNA
                     fragments derived from two cDNA clones (1 - 1066 was
                     derived from pf00933 and 1067 - 5189 was derived from
                     ha02567)."
     gene            1. .5189
                     /gene="KIAA0180"
     CDS             <1. .1213
                     /gene="KIAA0180"
                     /citation=[2]
                     /codon_start=2
                     /protein_id="BAA11497.2"
                     /db_xref="GI:20521838"
                     /translation="PPRLCAAAGSGSRADMGAAAWARPLSVSFLLLLLPLPGMPAGSW
                     DPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSY
                     QKYFKLEPLQAYHRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEG
                     NPFGPFWDQFHVSFNKSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLE
                     EHRPLQKYMVWSDEMVKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHF
                     MASPQCVGYSRSTAAPLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPEL
                     QQLFKGKVKVVSLKPEVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFG
                     MDRPPKLRDEF"
ORIGIN

Alignment Scores:
Pred. No.:              5.85e-35        Length:         5189
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     9               Gaps:           0

US-09-774-954-9 (1-61) x D80002 (1-5189)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 175

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 235

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 295

Qy         61 His 61
              |||
Db        296 CAT 298

RESULT 12
AF375884
LOCUS       AF375884                5249 bp    mRNA    linear   PRI 23-OCT-2001
DEFINITION  Homo sapiens protein o-fucosyltransferase (POFUT1) mRNA, complete
            cds.
ACCESSION   AF375884
VERSION     AF375884.1  GI:15825113
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 5249)
  AUTHORS   Wang,Y., Shao,L., Shi,S., Harris,R.J., Spellman,M.W., Stanley,P.
            and Haltiwanger,R.S.
  TITLE     Modification of epidermal growth factor-like repeats with O-fucose.
            Molecular cloning and expression of a novel GDP-fucose protein
            O-fucosyltransferase
  JOURNAL   J. Biol. Chem. 276 (43), 40338-40345 (2001)
  MEDLINE   21523965
   PUBMED   11524432
```

```
REFERENCE   2  (bases 1 to 5249)
  AUTHORS   Shi,S., Stanley,P., Wang,Y., Shao,L., Harris,R.J., Spellman,M.W.
            and Haltiwanger,R.S.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-MAY-2001) Cell Biology, Albert Einstein College of
            Medicine, 1300 Morris Park Ave., New York, NY 10461, USA
FEATURES             Location/Qualifiers
     source          1. .5249
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="between PLAGL2 and KIF3B"
                     /tissue_type="heart"
     gene            1. .5249
                     /gene="POFUT1"
     CDS             50. .1216
                     /gene="POFUT1"
                     /function="catalyzes transfer of fucose from GDP-fucose to
                     serine or threonine on epidermal growth factor-like
                     repeats"
                     /note="O-FucT-1; glycosyltransferase"
                     /codon_start=1
                     /product="protein o-fucosyltransferase"
                     /protein_id="AAL09576.1"
                     /db_xref="GI:15825114"
                     /translation="MGAAAWARPLSVSFLLLLLPLPGMPAGSWDPAGYLLYCPCMGRF
                     GNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLHVSYQKYFKLEPLQAYHRV
                     ISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHVSFN
                     KSELFTGISFSASYREQWSQRFSPKEHPVLALPGAPAQFPVLEEHRPLQKYMVWSDEM
                     VKTGEAQIHAHLVRPYVGIHLRIGSDWKNACAMLKDGTAGSHFMASPQCVGYSRSTAA
                     PLTMTMCLPDLKEIQRAVKLWVRSLDAQSVYVATDSESYVPELQQLFKGKVKVVSLKP
                     EVAQVDLYILGQADHFIGNCVSSFTAFVKRERDLQGRPSSFFGMDRPPKLRDEF"
ORIGIN

Alignment Scores:
Pred. No.:                5.92e-35        Length:         5249
Score:                    353.00          Matches:        61
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       9               Gaps:           0

US-09-774-954-9 (1-61) x AF375884 (1-5249)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy         61 His 61
              |||
Db        299 CAT 301

RESULT 13
AX780153
LOCUS       AX780153                5266 bp    DNA     linear   PAT 14-JUL-2003
DEFINITION  Sequence 2310 from Patent WO03039443.
ACCESSION   AX780153
VERSION     AX780153.1  GI:32697147
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Haferlach,T., Schoch,C., Kern,W., Kohlmann,A., Schnittger,S.,
            Dugas,M., Eils,R., Brors,B. and Mergenthaler,S.
  TITLE     Novel genetic markers for leukemias
  JOURNAL   Patent: WO 03039443-A 2310 15-MAY-2003;
            Deutsches Krebsforschungszentrum (DE) ;
            Ludwig-Maximilian-Universitaet Muenchen (DE) ; Haferlach, Torsten,
            PD Dr. Dr. (DE) ; Schoch, Claudia (DE) ; Kern, Wolfgang (DE)
FEATURES             Location/Qualifiers
     source          1. .5266
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Alignment Scores:
Pred. No.:                5.94e-35        Length:         5266
Score:                    353.00          Matches:        61
Percent Similarity:       100.00%         Conservative:   0
Best Local Similarity:    100.00%         Mismatches:     0
Query Match:              100.00%         Indels:         0
DB:                       6               Gaps:           0

US-09-774-954-9 (1-61) x AX780153 (1-5266)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 191

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 251

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 311

Qy         61 His 61
              |||
Db        312 CAT 314

RESULT 14
BD103190
LOCUS       BD103190                11284 bp    DNA     linear   PAT 27-AUG-2002
DEFINITION  O-fucosyltransferase.
ACCESSION   BD103190
VERSION     BD103190.1  GI:22648764
KEYWORDS    JP 2001527389-A/3.
SOURCE      unidentified
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 11284)
  AUTHORS   Wang,Y. and Spellman,M.W.
  TITLE     O-fucosyltransferase
  JOURNAL   Patent: JP 2001527389-A 3 25-DEC-2001;
            GENENTECH INC
COMMENT     OS   Unidentified
            PN   JP 2001527389-A/3
            PD   25-DEC-2001
            PF   17-DEC-1997 JP 1998532877
            PR   31-JAN-1997 US    08/792498,26-NOV-1997 US    08/978741 PI
            YANG WANG,MICHAEL W SPELLMAN
            PC   C12N15/54,C12N9/10,C07K16/40
            CC   Strandedness: Single;
            CC   Topology: Linear;
            CC   O-fucosyltransferase
            FH   Key             Location/Qualifiers
            FT   source          1. .11284
            FT                   /organism='Unidentified'.
FEATURES             Location/Qualifiers
     source          1. .11284
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
```

ORIGIN

Alignment Scores:

Pred. No.:	1.34e-34	Length:	11284
Score:	353.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-09-774-954-9 (1-61) x BD103190 (1-11284)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy         61 His 61
              |||
Db       4416 CAT 4418
```

RESULT 15
AY344581
LOCUS AY344581 1056 bp mRNA linear MAM 03-AUG-2003
DEFINITION Bos taurus protein O-fucosyltransferase 1b (pofut1) mRNA, complete cds.
ACCESSION AY344581
VERSION AY344581.1 GI:33303531
KEYWORDS .
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 1056)
AUTHORS Loriol,C., Germot,A., Dupuy,F. and Maftah,A.
TITLE Genomic organization and expression profile of O-fucosyltransferase genes, pofut1 and pofut2, in Bos taurus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1056)
AUTHORS Loriol,C.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2003) EA 3176 Glycobiologie et Biotechnologie, Faculte des Sciences et Techniques, 123, Avenue A. Thomas, Limoges 87060, France
FEATURES Location/Qualifiers
source 1. .1056
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
gene 1. .1056
/gene="pofut1"
CDS 1. .1056
/gene="pofut1"
/note="O-FucT-1b"
/codon_start=1
/product="protein O-fucosyltransferase 1b"
/protein_id="AAQ02333.1"
/db_xref="GI:33303532"
/translation="MGAAAWAPSLLPPRVSLLLLLPLPGLPVGSWDPAGYLLYCPCM
GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNVHVSYQKYFKLEPLQAY
HRVISLEDFMEKLAPTHWPPEKRVAYCFEVAAQRSPDKKTCPMKEGNPFGPFWDQFHV
SFNKSELFAGISFSASYKDQWIQRFSPEEHPVLALPGAPAQFPVLEEHRPLQKYMVWS
DEMVRTGEAQIHAHLIRPYVGIHLRIGSDWKNACAMLKDGTAGAHFMASPQCVGYSRH
TTAPLTMTMCLPDLKEIRRALKLWVTALNAQSVYIATDSESYLPEIQQLFKGKSLSFA
VPKASEMSLLLGDRKGK"

ORIGIN

Alignment Scores:

Pred. No.:	2.07e-34	Length:	1056
Score:	343.00	Matches:	58
Percent Similarity:	98.36%	Conservative:	2
Best Local Similarity:	95.08%	Mismatches:	1
Query Match:	97.17%	Indels:	0
DB:	4	Gaps:	0

US-09-774-954-9 (1-61) x AY344581 (1-1056)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              :::|||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 CTGCCCGTGGGCTCCTGGGATCCGGCCGGTTACCTGCTCTACTGCCCATGCATGGGGCGT 138

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 TTTGGGAACCAGGCTGATCATTTCTTGGGCTCCCTGGCATTTGCAAAGCTGTTGAACCGC 198

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        199 ACCCTGGCTGTACCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACGTC 258

Qy         61 His 61
              |||
Db        259 CAT 261
```

Search completed: October 26, 2005, 12:55:33
Job time : 1387.21 secs

against DNA

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 05:59:38 ; Search time 1065.43 Seconds
(without alignments)
2179.320 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105432_6406/app_query.fasta_1.917
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954_@CGN_1_1_5334_@runat_25102005_105432_6406 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	353	100.0	520	7	CR543310	CR543310 DKFZp459B
2	353	100.0	550	7	CN348927	CN348927 170005999
3	353	100.0	551	2	BE261136	BE261136 601152308
4	353	100.0	576	2	BE262956	BE262956 601147417
5	353	100.0	595	7	CN348926	CN348926 170005326
6	353	100.0	610	1	AL555719	AL555719 AL555719
7	353	100.0	656	7	CV029252	CV029252 7968 Full
8	353	100.0	699	1	AU137536	AU137536 AU137536
9	353	100.0	705	7	CN348925	CN348925 170005328
10	353	100.0	753	4	BG424731	BG424731 602453575
11	353	100.0	790	4	BG519639	BG519639 602578756
12	353	100.0	808	4	BG747760	BG747760 602705255
13	353	100.0	841	5	BX345989	BX345989 BX345989
14	353	100.0	889	4	BI524031	BI524031 603052202
15	353	100.0	898	2	BF315757	BF315757 601900828
16	353	100.0	919	5	BU155142	BU155142 AGENCOURT
17	353	100.0	942	7	CO774833	CO774833 ILLUMIGEN
18	353	100.0	950	4	BI199190	BI199190 602758739
19	353	100.0	955	2	BE260030	BE260030 601150556
20	353	100.0	1555	3	CR617950	CR617950 full-leng
21	353	100.0	1711	2	BF206548	BF206548 601869779
22	350	99.2	624	7	CN348922	CN348922 170004249
23	344	97.5	898	5	BX449172	BX449172 BX449172
24	343	97.2	417	2	AW415368	AW415368 49609 MAR
25	341	96.6	473	2	BF444498	BF444498 262391 MA
26	341	96.6	666	5	BP459416	BP459416 BP459416
27	341	96.6	736	5	BP457865	BP457865 BP457865
28	340	96.3	729	2	BE272703	BE272703 601105080
29	338	95.8	689	4	BI559635	BI559635 603252494
30	335	94.9	1241	4	BM478111	BM478111 AGENCOURT
31	333	94.3	874	4	BI757540	BI757540 603029693
32	331	93.8	318	6	BY778150	BY778150 BY778150
33	331	93.8	352	5	BY328132	BY328132 BY328132
34	331	93.8	456	2	BB633858	BB633858 BB633858
35	331	93.8	480	2	BB851550	BB851550 BB851550
36	331	93.8	536	7	CF155036	CF155036 B0620C09-
37	331	93.8	604	2	BB618795	BB618795 BB618795
38	331	93.8	605	6	BY728018	BY728018 BY728018
39	331	93.8	611	2	BB623905	BB623905 BB623905
40	331	93.8	614	6	BY752500	BY752500 BY752500
41	331	93.8	634	2	BB631131	BB631131 BB631131
42	331	93.8	641	2	BB611933	BB611933 BB611933
43	331	93.8	650	6	BY752497	BY752497 BY752497
44	331	93.8	661	2	BB642654	BB642654 BB642654
45	331	93.8	664	6	BY728337	BY728337 BY728337

ALIGNMENTS

RESULT 1
CR543310
LOCUS CR543310 520 bp mRNA linear EST 07-JUL-2004
DEFINITION DKFZp459B068_r1 459 (synonym: pcor1) Pongo pygmaeus cDNA clone DKFZp459B068 5', mRNA sequence.
ACCESSION CR543310
VERSION CR543310.1 GI:49894583
KEYWORDS EST.
SOURCE Pongo pygmaeus (orangutan)
ORGANISM Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE 1 (bases 1 to 520)
AUTHORS Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
TITLE Pongo pygmaeus mRNA (Koehrer,K., Beyer,A., Mewes,H.W., et al.)
JOURNAL Unpublished (2004)
COMMENT Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by BMFZ (Biomedical Research Center at the Heinrich-Heine-University, Duesseldorf/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp459B068) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.
FEATURES Location/Qualifiers
source 1. .520

```
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp459B068"
                     /tissue_type="cortex"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="459 (synonym: pcor1)"
                     /note="Vector: pSport1_Sfi; Site_1: SfilA; Site_2: SfilB"
ORIGIN

Alignment Scores:
Pred. No.:              1.52e-33        Length:         520
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-774-954-9 (1-61) x CR543310 (1-520)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        116 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 175

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        176 TTTGGGAACCAGGCGGATCACTTCTTGGGCTCTTTGGCATTTGCAAAGCTGCTAAACCGT 235

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        236 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCCTTCACCAACCTC 295

Qy         61 His 61
              |||
Db        296 CAT 298

RESULT 2
CN348927
LOCUS       CN348927                 550 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000599936929 GRN_PRENEU Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN348927
VERSION     CN348927.1  GI:47348861
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 550)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 550   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .550
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cell, retinoic acid and
                     mitogen-treated hES cell line H7"
                     /clone_lib="GRN_PRENEU"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from hES cell line H7 (p29) maintained in feeder-free
                     conditions. Embryoid bodies were generated in the presence
                     of all-trans retinoic and mitogens."
ORIGIN

Alignment Scores:
Pred. No.:              1.63e-33        Length:         550
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-774-954-9 (1-61) x CN348927 (1-550)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        109 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 168

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        169 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 228

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 288

Qy         61 His 61
              |||
Db        289 CAT 291

RESULT 3
BE261136
LOCUS       BE261136                 551 bp    mRNA    linear   EST 26-OCT-2000
DEFINITION  601152308F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3508464 5',
            mRNA sequence.
ACCESSION   BE261136
VERSION     BE261136.1  GI:9132941
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 551)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM189  row: p  column: 01
            High quality sequence stop: 551.
FEATURES             Location/Qualifiers
     source          1. .551
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3508464"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_19"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
```

ORIGIN

Alignment Scores:

Pred. No.:	1.64e-33	Length:	551
Score:	353.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-774-954-9 (1-61) x BE261136 (1-551)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 159

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 219

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 279

Qy         61 His 61
              |||
Db        280 CAT 282
```

RESULT 4
BE262956

```
LOCUS       BE262956                576 bp    mRNA    linear   EST 26-OCT-2000
DEFINITION  601147417F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3162533 5',
            mRNA sequence.
ACCESSION   BE262956
VERSION     BE262956.1  GI:9136425
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM122  row: b  column: 06
            High quality sequence start: 11
            High quality sequence stop: 576.
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3162533"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_19"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN
```

Alignment Scores:

Pred. No.:	1.73e-33	Length:	576
Score:	353.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-774-954-9 (1-61) x BE262956 (1-576)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        154 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 213

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        214 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 273

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        274 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 333

Qy         61 His 61
              |||
Db        334 CAT 336
```

RESULT 5
CN348926

```
LOCUS       CN348926                595 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000532646873 GRN_ES Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN348926
VERSION     CN348926.1  GI:47348860
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 595)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 595   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .595
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, cell lines H1, H7, and
                     H9"
                     /clone_lib="GRN_ES"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from undifferentiated hES cell lines H1 (p32), H7 (p29),
                     and H9 (p26) maintained in feeder-free conditions"
ORIGIN
```

Alignment Scores:

Pred. No.:	1.8e-33	Length:	595
Score:	353.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	7	Gaps:	0

```
US-09-774-954-9 (1-61) x CN348926 (1-595)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         99 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 158

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        159 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 218

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        219 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 278

Qy         61 His 61
              |||
Db        279 CAT 281

RESULT 6
AL555719
LOCUS       AL555719                610 bp    mRNA    linear   EST 30-MAR-2004
DEFINITION  AL555719 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DK003YN13 5-PRIME, mRNA sequence.
ACCESSION   AL555719
VERSION     AL555719.3  GI:45860448
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 15, 2001 this sequence version replaced gi:31277525.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            9980.f
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0DK003CG07QP1&c=9980.f.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DK003YN13"
                     /cell_type="HELA CELLS COT 25-NORMALIZED"
                     /cell_line="HELA"
                     /clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Alignment Scores:
Pred. No.:              1.85e-33        Length:         610
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-774-954-9 (1-61) x AL555719 (1-610)
```

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        113 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 172

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        173 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 232

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        233 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 292

Qy         61 His 61
              |||
Db        293 CAT 295

RESULT 7
CV029252
LOCUS       CV029252                656 bp    mRNA    linear   EST 20-AUG-2004
DEFINITION  7968 Full Length cDNA from the Mammalian Gene Collection Homo
            sapiens cDNA 5' similar to BC000582, mRNA sequence.
ACCESSION   CV029252
VERSION     CV029252.1  GI:51487402
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 656)
  AUTHORS   Rual,J.F., Hirozane-Kishikawa,T., Hao,T., Bertin,N., Li,S.,
            Dricot,A., Li,N., Rosenberg,J., Lamesch,P., Vidalain,P.O.,
            Clingingsmith,T.R., Hartley,J.L., Esposito,D., Cheo,D., Moore,T.,
            Simmons,B., Sequerra,R., Bosak,S., Doucette-Stamm,L., Le Peuch,C.,
            Vandenhaute,J., Cusick,M.E., Albala,J.S., Hill,D.E. and Vidal,M.
  TITLE     Human ORFeome Version 1.1: a Platform for Reverse Proteomics
  JOURNAL   Genome Res. (2004) In press
COMMENT     Contact: Vidal M
            Marc Vidal Laboratory
            Dana Farber Cancer Institute
            1 Jimmy Fund Way Smith 858, BOSTON, MA 02115, USA
            Tel: 617 632 5180
            Fax: 617 632 5739
            Email: Marc_Vidal@dfci.harvard.edu
            ORF Sequence Tag (OST) of Gateway Entry construct. Each cloned ORF
            results from a PCR reaction using an MGC full-length cDNA as
            template DNA and ORF specific primers
            PCR PRimers
            FORWARD: ATGGGCGCCGCCGCGTGGGC
            BACKWARD: CACCTTGGGAAGAGTAAGGTA
            Insert Length: 656    Std Error: 18.00
            Plate: 11019  row: 04  column: B
            Seq primer: ACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC
            High quality sequence start: 87
            High quality sequence stop: 655
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="mixed"
                     /clone_lib="Full Length cDNA from the Mammalian Gene
                     Collection"
                     /note="Vector: mixed; The ORFs were PCR amplified from the
                     MGC (Mammalian Gene Collection) as of April 2004 and
                     cloned by recombinational Gateway cloning into pDONR223
                     Donor vector. Reference : MGC (Mammalian Gene Collection)
                     Program Team, Generation and Initial Analysis of more than
                     15,000 Full-Length Human and Mouse cDNA Sequences. PNAS,
                     2002, 99(26), 16899-16903"
ORIGIN

Alignment Scores:
```

```
Pred. No.:              2.02e-33        Length:         656
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-774-954-9 (1-61) x CV029252 (1-656)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         70 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 129

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        130 TTTGGGAAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 189

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 249

Qy         61 His 61
              |||
Db        250 CAT 252

RESULT 8
AU137536
LOCUS       AU137536                699 bp    mRNA    linear   EST 02-AUG-2002
DEFINITION  AU137536 PLACE1 Homo sapiens cDNA clone PLACE1006691 5', mRNA
            sequence.
ACCESSION   AU137536
VERSION     AU137536.1  GI:10998075
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 699)
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
            Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .699
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="PLACE1006691"
                     /tissue_type="placenta"
                     /clone_lib="PLACE1"
                     /note="Vector: pME18SFL3"
ORIGIN

Alignment Scores:
Pred. No.:              2.18e-33        Length:         699
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     1               Gaps:           0

US-09-774-954-9 (1-61) x AU137536 (1-699)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 191

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 251

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 311

Qy         61 His 61
              |||
Db        312 CAT 314

RESULT 9
CN348925
LOCUS       CN348925                705 bp    mRNA    linear   EST 16-MAY-2004
DEFINITION  17000532814105 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN348925
VERSION     CN348925.1  GI:47348859
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 705)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE     Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT     Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 705   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .705
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, embryoid bodies
                     derived from H1, H7 and H9 cells"
                     /clone_lib="GRN_EB"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from embryoid body outgrowths derived from hES cell lines
                     H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
                     conditions."
ORIGIN

Alignment Scores:
Pred. No.:              2.21e-33        Length:         705
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     7               Gaps:           0

US-09-774-954-9 (1-61) x CN348925 (1-705)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         50 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 109

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        110 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 169

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 229

Qy         61 His 61
                |||
Db        230 CAT 232

RESULT 10
BG424731
LOCUS       BG424731                753 bp    mRNA    linear   EST 14-MAR-2001
DEFINITION  602453575F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4591867 5',
            mRNA sequence.
ACCESSION   BG424731
VERSION     BG424731.1  GI:13331237
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 753)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1329  row: e  column: 20
            High quality sequence stop: 699.
FEATURES             Location/Qualifiers
     source          1. .753
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4591867"
                     /tissue_type="renal cell adenocarcinoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_14"
                     /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

Alignment Scores:
Pred. No.:              2.39e-33        Length:         753
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     4               Gaps:           0

US-09-774-954-9 (1-61) x BG424731 (1-753)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 180

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 240

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 300

Qy         61 His 61
                |||
Db        301 CAT 303

RESULT 11
BG519639
LOCUS       BG519639                790 bp    mRNA    linear   EST 02-APR-2001
DEFINITION  602578756F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503461 5',
            mRNA sequence.
ACCESSION   BG519639
VERSION     BG519639.1  GI:13515377
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 790)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM176  row: o  column: 14
            High quality sequence stop: 710.
FEATURES             Location/Qualifiers
     source          1. .790
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3503461"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_19"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:              2.53e-33        Length:         790
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     4               Gaps:           0

US-09-774-954-9 (1-61) x BG519639 (1-790)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 179

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 239
```

```
Qy          41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         240 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 299

Qy          61 His 61
               |||
Db         300 CAT 302

RESULT 12
BG747760
LOCUS       BG747760                808 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602705255F1 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:4842061 5',
            mRNA sequence.
ACCESSION   BG747760
VERSION     BG747760.1  GI:14058413
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 808)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1675  row: n  column: 14
            High quality sequence stop: 808.
FEATURES             Location/Qualifiers
     source          1. .808
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4842061"
                     /tissue_type="normal pigmented retinal epithelium"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_43"
                     /note="Organ: eye; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library. |"
ORIGIN

Alignment Scores:
Pred. No.:              2.6e-33         Length:         808
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     4               Gaps:           0

US-09-774-954-9 (1-61) x BG747760 (1-808)

Qy           1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          82 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 141

Qy          21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         142 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 201

Qy          41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         202 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 261

Qy          61 His 61
               |||
Db         262 CAT 264

RESULT 13
BX345989
LOCUS       BX345989                841 bp    mRNA    linear   EST 23-APR-2004
DEFINITION  BX345989 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED
            Homo sapiens cDNA clone CS0DJ008YF12 5-PRIME, mRNA sequence.
ACCESSION   BX345989
VERSION     BX345989.2  GI:46549776
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 841)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On May 5, 2003 this sequence version replaced gi:30377055.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            9980.f
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0AJ008DC06QP1&c=9980.f.
FEATURES             Location/Qualifiers
     source          1. .841
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DJ008YF12"
                     /cell_type="T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED"
                     /cell_line="JURKAT"
                     /clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE) COT
                     10-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Alignment Scores:
Pred. No.:              2.73e-33        Length:         841
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     5               Gaps:           0

US-09-774-954-9 (1-61) x BX345989 (1-841)

Qy           1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         109 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 168

Qy          21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         169 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 228

Qy          41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         229 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 288
```

```
Qy          61 His 61
               |||
Db         289 CAT 291

RESULT 14
BI524031
LOCUS       BI524031                889 bp    mRNA    linear   EST 29-AUG-2001
DEFINITION  603052202F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5201616 5',
            mRNA sequence.
ACCESSION   BI524031
VERSION     BI524031.1  GI:15348823
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 889)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11505  row: d  column: 01
            High quality sequence stop: 880.
FEATURES             Location/Qualifiers
     source          1. .889
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5201616"
                     /lab_host="DH10B"
                     /clone_lib="NIH_MGC_122"
                     /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
                     anonymous pool of 24 week female lung, 16 week female
                     spleen, and 20-22 week male spleens.  Library is oligo-dT
                     primed and directionally cloned (EcoRV site is destroyed
                     upon  cloning).  Average insert size 1.4 kb, insert size
                     range 1-3 kb. Library is normalized and enriched for
                     full-length clones and was  constructed by C. Gruber
                     (Invitrogen).  Research Genetics tracking  code 026. Note:
                     this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:              2.92e-33        Length:         889
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     4               Gaps:           0

US-09-774-954-9 (1-61) x BI524031 (1-889)

Qy           1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          87 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 146

Qy          21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         147 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 206

Qy          41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         207 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 266

Qy          61 His 61
               |||
Db         267 CAT 269

RESULT 15
BF315757
LOCUS       BF315757                898 bp    mRNA    linear   EST 21-NOV-2000
DEFINITION  601900828F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:4129975 5',
            mRNA sequence.
ACCESSION   BF315757
VERSION     BF315757.1  GI:11264027
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 898)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM1025  row: h  column: 08
            High quality sequence stop: 721.
FEATURES             Location/Qualifiers
     source          1. .898
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4129975"
                     /tissue_type="neuroblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_19"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Library constructed  by Ling Hong
                     in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and  Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:              2.96e-33        Length:         898
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     2               Gaps:           0

US-09-774-954-9 (1-61) x BF315757 (1-898)

Qy           1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         120 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 179

Qy          21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         180 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 239

Qy          41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         240 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 299

Qy          61 His 61
```

Db 300 CAT 302

Search completed: October 26, 2005, 15:17:27
Job time : 1069.43 secs

against DNA

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 09:31:14 ; Search time 260.659 Seconds
(without alignments)
1931.299 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 9772377 seqs, 4126317084 residues

Total number of hits satisfying chosen parameters: 19544754

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105434_6433/app_query.fasta_1.917
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09774954_@CGN_1_1_684_@runat_25102005_105434_6433
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
19: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
20: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
21: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
22: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
23: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
24: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq2:*
25: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
26: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
27: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
28: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	353	100.0	1100	11	US-09-774-954-16	Sequence 16, Appl
	2	353	100.0	1300	11	US-09-774-954-4	Sequence 4, Appli
	3	353	100.0	1400	22	US-10-956-157-9986	Sequence 9986, Ap
	4	353	100.0	1506	22	US-10-956-157-4751	Sequence 4751, Ap
	5	353	100.0	1514	11	US-09-774-954-1	Sequence 1, Appli
	6	353	100.0	5218	16	US-10-301-822-168	Sequence 168, App
	7	353	100.0	5218	22	US-10-956-157-4752	Sequence 4752, Ap
	8	353	100.0	11284	11	US-09-774-954-5	Sequence 5, Appli
	9	223	63.2	1209	26	US-11-097-143-18161	Sequence 18161, A
	10	218	61.8	5009	11	US-09-774-954-7	Sequence 7, Appli
c	11	188	53.3	3264	26	US-11-097-143-18160	Sequence 18160, A
c	12	188	53.3	3793	26	US-11-097-143-18082	Sequence 18082, A
c	13	120	34.0	521	13	US-09-925-065A-602444	Sequence 602444,
	14	120	34.0	760	24	US-10-820-474A-236	Sequence 236, App
	15	90	25.5	591	9	US-09-864-761-9418	Sequence 9418, Ap
	16	90	25.5	591	9	US-09-864-761-26446	Sequence 26446, A
	17	90	25.5	591	17	US-10-029-386-24815	Sequence 24815, A
	18	86	24.4	447	21	US-10-425-115-142368	Sequence 142368,
	19	82	23.2	402	9	US-09-783-590-11501	Sequence 11501, A
	20	80.5	22.8	829	19	US-10-424-599-123276	Sequence 123276,
	21	77.5	22.0	568	21	US-10-425-115-16666	Sequence 16666, A
	22	77.5	22.0	1617	20	US-10-437-963-11436	Sequence 11436, A
c	23	75.5	21.4	277616	20	US-10-367-094-83	Sequence 83, Appl
	24	73.5	20.8	1803	16	US-10-156-761-5001	Sequence 5001, Ap
	25	73.5	20.8	9025608	16	US-10-156-761-1	Sequence 1, Appli
	26	73	20.7	3545	26	US-11-097-143-11746	Sequence 11746, A
c	27	72	20.4	430	9	US-09-960-352-10634	Sequence 10634, A
c	28	72	20.4	432	9	US-09-960-352-1656	Sequence 1656, Ap
c	29	72	20.4	447	9	US-09-960-352-6689	Sequence 6689, Ap
	30	71	20.1	1186	18	US-10-282-122A-26446	Sequence 26446, A
	31	71	20.1	1242	18	US-10-282-122A-28170	Sequence 28170, A
c	32	71	20.1	4953	20	US-10-437-963-23967	Sequence 23967, A
	33	70.5	20.0	467	10	US-09-918-995-29055	Sequence 29055, A
c	34	70.5	20.0	510	17	US-10-029-386-10060	Sequence 10060, A
c	35	70	19.8	326002	22	US-10-461-862-56	Sequence 56, Appl
c	36	69.5	19.7	622	20	US-10-437-963-15591	Sequence 15591, A
c	37	69	19.5	600	24	US-10-972-079-32895	Sequence 32895, A
c	38	68.5	19.4	600	24	US-10-972-079-43878	Sequence 43878, A
c	39	68.5	19.4	600	24	US-10-972-079-43879	Sequence 43879, A
c	40	68.5	19.4	600	24	US-10-972-079-43880	Sequence 43880, A
	41	68.5	19.4	3373	15	US-10-105-963-9	Sequence 9, Appli
	42	68.5	19.4	8174	9	US-09-863-475A-5	Sequence 5, Appli
c	43	67.5	19.1	358	21	US-10-425-115-50448	Sequence 50448, A
	44	67.5	19.1	600	24	US-10-972-079-27372	Sequence 27372, A
c	45	67.5	19.1	636	13	US-09-925-065A-285729	Sequence 285729,

ALIGNMENTS

RESULT 1
US-09-774-954-16
; Sequence 16, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA

```
;            ZIP: 94080
;        COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/774,954
;            FILING DATE: 30-Jan-2001
;            CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/978,741
;            FILING DATE: 26-NOV-1997
;            APPLICATION NUMBER: 08/792,498
;            FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 16:
;        SEQUENCE CHARACTERISTICS:
;            LENGTH: 1100 base pairs
;            TYPE: Nucleic Acid
;            STRANDEDNESS: Single
;            TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-774-954-16

Alignment Scores:
Pred. No.:              5.82e-40        Length:         1100
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     11              Gaps:           0

US-09-774-954-9 (1-61) x US-09-774-954-16 (1-1100)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 His 61
              |||
Db        181 CAT 183

RESULT 2
US-09-774-954-4
; Sequence 4, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genentech, Inc.
;            STREET: 1 DNA Way
;            CITY: South San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94080
;        COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/774,954
;            FILING DATE: 30-Jan-2001
;            CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/978,741
;            FILING DATE: 26-NOV-1997
;            APPLICATION NUMBER: 08/792,498
;            FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;            LENGTH: 1300 base pairs
;            TYPE: Nucleic Acid
;            STRANDEDNESS: Single
;            TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-774-954-4

Alignment Scores:
Pred. No.:              7.1e-40         Length:         1300
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     11              Gaps:           0

US-09-774-954-9 (1-61) x US-09-774-954-4 (1-1300)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy         61 His 61
              |||
Db        316 CAT 318

RESULT 3
US-10-956-157-9986
; Sequence 9986, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William
;  TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;  TITLE OF INVENTION:  HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;  FILE REFERENCE: 031896-043000 (AM 101081)
;  CURRENT APPLICATION NUMBER: US/10/956,157
;  CURRENT FILING DATE:  2004-10-04
;  NUMBER OF SEQ ID NOS: 319805
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9986
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

US-10-956-157-9986

Alignment Scores:

Pred. No.:	7.75e-40	Length:	1400
Score:	353.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0

US-09-774-954-9 (1-61) x US-10-956-157-9986 (1-1400)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         13 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 72

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         73 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 132

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 192

Qy         61 His 61
              |||
Db        193 CAT 195
```

RESULT 4
US-10-956-157-4751
; Sequence 4751, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4751
; LENGTH: 1506
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-956-157-4751

Alignment Scores:

Pred. No.:	8.45e-40	Length:	1506
Score:	353.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	22	Gaps:	0

US-09-774-954-9 (1-61) x US-10-956-157-4751 (1-1506)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy         61 His 61
              |||
Db        299 CAT 301
```

RESULT 5
US-09-774-954-1
; Sequence 1, Application US/09774954
; Publication No. US20040241645A1
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/774,954
; FILING DATE: 30-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-NOV-1997
; APPLICATION NUMBER: 08/792,498
; FILING DATE: 31-JAN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1514 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-774-954-1

Alignment Scores:

Pred. No.:	8.51e-40	Length:	1514
Score:	353.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	11	Gaps:	0

US-09-774-954-9 (1-61) x US-09-774-954-1 (1-1514)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 His 61
              |||
Db        181 CAT 183
```

RESULT 6

```
US-10-301-822-168
; Sequence 168, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
;  APPLICANT:  Burgart, Lawrence J.
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
;  CURRENT APPLICATION NUMBER: US/10/301,822
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 168
;   LENGTH: 5218
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (50)...(1216)
US-10-301-822-168

Alignment Scores:
Pred. No.:              3.7e-39       Length:         5218
Score:                  353.00        Matches:        61
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     16            Gaps:           0

US-09-774-954-9 (1-61) x US-10-301-822-168 (1-5218)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy         61 His 61
              |||
Db        299 CAT 301

RESULT 7
US-10-956-157-4752
; Sequence 4752, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:
;  APPLICANT: Wyeth
;  APPLICANT:  Mounts, William
;  TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
;  TITLE OF INVENTION:  HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
;  FILE REFERENCE: 031896-043000 (AM 101081)
;  CURRENT APPLICATION NUMBER: US/10/956,157
;  CURRENT FILING DATE:  2004-10-04
;  NUMBER OF SEQ ID NOS: 319805
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4752
;   LENGTH: 5218
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-956-157-4752

Alignment Scores:
Pred. No.:              3.7e-39       Length:         5218
Score:                  353.00        Matches:        61
Percent Similarity:     100.00%       Conservative:   0
Best Local Similarity:  100.00%       Mismatches:     0
Query Match:            100.00%       Indels:         0
DB:                     22            Gaps:           0

US-09-774-954-9 (1-61) x US-10-956-157-4752 (1-5218)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy         61 His 61
              |||
Db        299 CAT 301

RESULT 8
US-09-774-954-5
; Sequence 5, Application US/09774954
; Publication No. US20040241645A1
;  GENERAL INFORMATION:
;       APPLICANT: Yang Wang, Michael W. Spellman
;       TITLE OF INVENTION: O-Fucosyltransferase
;       NUMBER OF SEQUENCES: 17
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genentech, Inc.
;            STREET: 1 DNA Way
;            CITY: South San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94080
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: WinPatin (Genentech)
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/774,954
;            FILING DATE: 30-Jan-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/978,741
;            FILING DATE: 26-NOV-1997
;            APPLICATION NUMBER: 08/792,498
;            FILING DATE: 31-JAN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 5:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 11284 base pairs
```

```
;             TYPE: Nucleic Acid
;             STRANDEDNESS: Single
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-774-954-5

Alignment Scores:
Pred. No.:              9.25e-39        Length:         11284
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     11              Gaps:           0

US-09-774-954-9 (1-61) x US-09-774-954-5 (1-11284)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy         61 His 61
              |||
Db       4416 CAT 4418

RESULT 9
US-11-097-143-18161
; Sequence 18161, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18161
;   LENGTH: 1209
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-18161

Alignment Scores:
Pred. No.:              2.25e-21        Length:         1209
Score:                  223.00          Matches:        38
Percent Similarity:     91.11%          Conservative:   3
Best Local Similarity:  84.44%          Mismatches:     4
Query Match:            63.17%          Indels:         0
DB:                     26              Gaps:           0

US-09-774-954-9 (1-61) x US-11-097-143-18161 (1-1209)

Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGlnAlaAsp 26
              ||||||  |||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db         82 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCGAC 141

Qy         27 HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProPro 46
              |||||||||||||||||||||||||||||  ||||||||||||||  :::||||||
Db        142 CACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCG 201

Qy         47 TrpIleGluTyrGln 51
              |||:::||||||:::
Db        202 TGGGTGGAGTATCGT 216

RESULT 10
US-09-774-954-7
; Sequence 7, Application US/09774954
; Publication No. US20040241645A1
;  GENERAL INFORMATION:
;       APPLICANT: Yang Wang, Michael W. Spellman
;       TITLE OF INVENTION: O-Fucosyltransferase
;       NUMBER OF SEQUENCES: 17
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genentech, Inc.
;            STREET: 1 DNA Way
;            CITY: South San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94080
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: WinPatin (Genentech)
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/774,954
;            FILING DATE: 30-Jan-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/978,741
;            FILING DATE: 26-NOV-1997
;            APPLICATION NUMBER: 08/792,498
;            FILING DATE: 31-JAN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;  INFORMATION FOR SEQ ID NO: 7:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 5009 base pairs
;            TYPE: Nucleic Acid
;            STRANDEDNESS: Single
;            TOPOLOGY: Linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-774-954-7

Alignment Scores:
Pred. No.:              6.29e-20        Length:         5009
Score:                  218.00          Matches:        39
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            61.76%          Indels:         0
DB:                     11              Gaps:           0

US-09-774-954-9 (1-61) x US-09-774-954-7 (1-5009)
```

```
Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 61

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 118

RESULT 11
US-11-097-143-18160/c
; Sequence 18160, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18160
;   LENGTH: 3264
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-18160

Alignment Scores:
Pred. No.:              7.17e-16        Length:         3264
Score:                  188.00          Matches:        38
Percent Similarity:     64.06%          Conservative:   3
Best Local Similarity:  59.38%          Mismatches:     4
Query Match:            53.26%          Indels:         19
DB:                     26              Gaps:           1

US-09-774-954-9 (1-61) x US-11-097-143-18160 (1-3264)

Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet------------------------ 18
              ||||||   |||||||||   ||||||||||||||||
Db       2183 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGTAAGTTTCACTTGGAAGCAACC 2124

Qy         19 ------------------------------GlyArgPheGlyAsnGlnAlaAspHisPh 28
                                            ||||||||||||||||||||||||||||
Db       2123 GAAAAACGACAACTAAACCGCGCTTTTCAAAGGACGCTTTGGCAACCAGGCCGACCACTT 2064

Qy         28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProProTrpIl 48
              ||||||||||||||||||||||   ||||||||||||||   :::|||||||||::
Db       2063 CCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCGTGGGT 2004

Qy         48 eGluTyrGln 51
              :||||||:::
Db       2003 GGAGTATCGT 1994

RESULT 12
US-11-097-143-18082/c
; Sequence 18082, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18082
;   LENGTH: 3793
;   TYPE: DNA
;   ORGANISM: DROSOPHILA
US-11-097-143-18082

Alignment Scores:
Pred. No.:              8.57e-16        Length:         3793
Score:                  188.00          Matches:        38
Percent Similarity:     64.06%          Conservative:   3
Best Local Similarity:  59.38%          Mismatches:     4
Query Match:            53.26%          Indels:         19
DB:                     26              Gaps:           1

US-09-774-954-9 (1-61) x US-11-097-143-18082 (1-3793)

Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet------------------------ 18
              ||||||   |||||||||   ||||||||||||||||
Db        675 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGTAAGTTTCACTTGGAAGCAACC 616

Qy         19 ------------------------------GlyArgPheGlyAsnGlnAlaAspHisPh 28
                                            ||||||||||||||||||||||||||||
Db        615 GAAAAACGACAACTAAACCGCGCTTTTCAAAGGACGCTTTGGCAACCAGGCCGACCACTT 556

Qy         28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProProTrpIl 48
              ||||||||||||||||||||||   ||||||||||||||   :::|||||||||::
Db        555 CCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCGTGGGT 496

Qy         48 eGluTyrGln 51
              :||||||:::
Db        495 GGAGTATCGT 486

RESULT 13
US-09-925-065A-602444/c
; Sequence 602444, Application US/09925065A
; Publication No. US20050228172A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single
;  TITLE OF INVENTION:  Nucleotide Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.135
;  CURRENT APPLICATION NUMBER: US/09/925,065A
;  CURRENT FILING DATE:  2001-08-08
;  PRIOR APPLICATION NUMBER: US 60/243,096
```

```
;  PRIOR FILING DATE: 2000-10-24
;  PRIOR APPLICATION NUMBER: US 60/252,147
;  PRIOR FILING DATE: 2000-11-20
;  PRIOR APPLICATION NUMBER: US 60/250,092
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: US 60/261,766
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/289,846
;  PRIOR FILING DATE: 2001-05-09
;  NUMBER OF SEQ ID NOS: 957086
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 602444
;   LENGTH: 521
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-925-065A-602444

Alignment Scores:
Pred. No.:                4.03e-07      Length:          521
Score:                    120.00        Matches:         19
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    95.00%        Mismatches:      0
Query Match:              33.99%        Indels:          0
DB:                       13            Gaps:            0

US-09-774-954-9 (1-61) x US-09-925-065A-602444 (1-521)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        286 ATGCCYGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGTAAG 227

RESULT 14
US-10-820-474A-236
; Sequence 236, Application US/10820474A
; Publication No. US20050155089A1
; GENERAL INFORMATION:
;  APPLICANT: LAL, PREETI
;  APPLICANT:  TANG, Y. TOM
;  APPLICANT:  GORGONE, GINA A.
;  APPLICANT:  CORLEY, NEIL C.
;  APPLICANT:  GUEGLER, KARL J.
;  APPLICANT:  BAUGHN, MARIAH R.
;  APPLICANT:  AKERBLOM, INGRID E.
;  APPLICANT:  AU-YOUNG, JANICE
;  APPLICANT:  YUE, HENRY
;  APPLICANT:  PATTERSON, CHANDRA
;  APPLICANT:  REDDY, ROOPA
;  APPLICANT:  HILLMAN, JENNIFER L.
;  APPLICANT:  BANDMAN, OLGA
;  TITLE OF INVENTION: SIGNAL PEPTIDE-CONTAINING MOLECULES
;  FILE REFERENCE: 039386-1568
;  CURRENT APPLICATION NUMBER: US/10/820,474A
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: 09/720,533
;  PRIOR FILING DATE: 2001-03-20
;  PRIOR APPLICATION NUMBER: PCT/US99/14484
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/090,762
;  PRIOR FILING DATE: 1998-06-26
;  PRIOR APPLICATION NUMBER: 60/094,983
;  PRIOR FILING DATE: 1998-07-31
;  PRIOR APPLICATION NUMBER: 60/102,686
;  PRIOR FILING DATE: 1998-10-01
;  NUMBER OF SEQ ID NOS: 269
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 236
;   LENGTH: 760
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte Clone No: 1647884
US-10-820-474A-236

Alignment Scores:
Pred. No.:                6.32e-07      Length:          760
Score:                    120.00        Matches:         19
Percent Similarity:       100.00%       Conservative:    1
Best Local Similarity:    95.00%        Mismatches:      0
Query Match:              33.99%        Indels:          0
DB:                       24            Gaps:            0

US-09-774-954-9 (1-61) x US-10-820-474A-236 (1-760)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        111 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGTAAG 170

RESULT 15
US-09-864-761-9418
; Sequence 9418, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9418
;   LENGTH: 591
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC004624.6
```

```
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL =   2
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.3
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.2
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
;   OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 4.4
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5
US-09-864-761-9418

Alignment Scores:
Pred. No.:              0.00889         Length:         591
Score:                  90.00           Matches:        21
Percent Similarity:     41.18%          Conservative:   7
Best Local Similarity:  30.88%          Mismatches:     18
Query Match:            25.50%          Indels:         22
DB:                     9               Gaps:           2

US-09-774-954-9 (1-61) x US-09-864-761-9418 (1-591)

Qy          2 ProAlaGlySerTrpAspProAlaGlyTyrLeu---LeuTyrCysProCysMetGlyArg 20
              |||   ||||||||||||||||||||   |||   |||:::     |||
Db        387 CCGCAGGGCAGTTGGGACCCCGCGGGCCTGCTCAGACTACACGGGTCGTGCCCAGCGCCA 446

Qy         21 PheGlyAsnGlnAlaAsp---------------------------------------- 26
              ||||||:::   :::|||
Db        447 TTTGGGCACTCCAGTGACCAGGGCCTCGGGTCCTGCACACCTCCCATGAGGAGAGATGTC 506

Qy         27 --------------------HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsn 39
                                  |||      ||||||:::   :::   ::::::
Db        507 TTCCTCACATGGAATCAAAGCCACATCTGTGGTTCTATCCTCTGGGACCAGGTCTGCAGA 566

Qy         40 ArgThrLeuAlaValProProTrp 47
                |||      |||||||||
Db        567 GCTGGTTTAGGTCATCCTCCTTGG 590


Search completed: October 26, 2005, 16:00:15
Job time : 272.326 secs
```

Against DNA

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 08:18:39 ; Search time 53.4846 Seconds
(without alignments)
1866.199 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105433_6415/app_query.fasta_1.917
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954_@CGN_1_1_116_@runat_25102005_105433_6415 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	353	100.0	1100	3	US-08-978-741-16	Sequence 16, Appl
2	353	100.0	1100	3	US-09-333-729A-16	Sequence 16, Appl
3	353	100.0	1300	3	US-08-978-741-4	Sequence 4, Appli
4	353	100.0	1300	3	US-09-333-729A-6	Sequence 6, Appli
5	353	100.0	1514	3	US-08-978-741-1	Sequence 1, Appli
6	353	100.0	1514	3	US-09-333-729A-2	Sequence 2, Appli
7	353	100.0	11284	3	US-08-978-741-5	Sequence 5, Appli
8	223	63.2	1320	4	US-09-270-767-14353	Sequence 14353, A
9	218	61.8	5009	3	US-08-978-741-7	Sequence 7, Appli
10	218	61.8	5009	3	US-09-333-729A-8	Sequence 8, Appli
c 11	120	34.0	19227	4	US-09-949-016-12127	Sequence 12127, A
c 12	120	34.0	19228	4	US-09-949-016-16285	Sequence 16285, A
c 13	75.5	21.4	260247	4	US-09-949-016-13358	Sequence 13358, A
14	71	20.1	4403765	3	US-09-103-840A-2	Sequence 2, Appli
15	71	20.1	4411529	3	US-09-103-840A-1	Sequence 1, Appli
16	70.5	20.0	36759	4	US-09-949-016-12216	Sequence 12216, A
17	70.5	20.0	36760	4	US-09-949-016-14021	Sequence 14021, A
18	68.5	19.4	999	5	PCT-US91-00899-10	Sequence 10, Appl
19	68.5	19.4	1136	1	US-08-395-800A-9	Sequence 9, Appli
20	68.5	19.4	1155	1	US-08-434-151-1	Sequence 1, Appli
21	68.5	19.4	1155	1	US-08-208-889A-1	Sequence 1, Appli
22	68.5	19.4	1155	2	US-08-433-271-1	Sequence 1, Appli
23	68.5	19.4	1155	2	US-08-715-259-1	Sequence 1, Appli
24	68.5	19.4	1174	5	PCT-US95-07554-3	Sequence 3, Appli
25	68.5	19.4	1199	1	US-08-395-800A-5	Sequence 5, Appli
26	68.5	19.4	2268	3	US-08-675-773B-4	Sequence 4, Appli
27	68.5	19.4	3373	1	US-08-273-411-2	Sequence 2, Appli
28	68.5	19.4	3374	4	US-09-949-016-5482	Sequence 5482, Ap
29	68.5	19.4	3791	3	US-08-675-773B-3	Sequence 3, Appli
30	68.5	19.4	8174	1	US-07-914-281-5	Sequence 5, Appli
31	68.5	19.4	8174	1	US-08-393-246-5	Sequence 5, Appli
32	68.5	19.4	8174	1	US-08-525-058A-5	Sequence 5, Appli
33	68.5	19.4	8174	2	US-08-696-731-5	Sequence 5, Appli
34	68.5	19.4	8174	3	US-09-042-531-5	Sequence 5, Appli
35	68.5	19.4	8174	5	PCT-US91-00899-3	Sequence 3, Appli
36	68.5	19.4	8580	4	US-09-949-016-17224	Sequence 17224, A
37	66.5	18.8	26896	4	US-09-949-016-16800	Sequence 16800, A
38	66	18.7	1029	4	US-09-244-805-60	Sequence 60, Appl
39	65.5	18.6	87644	4	US-09-949-016-16041	Sequence 16041, A
40	65	18.4	601	4	US-09-949-016-84839	Sequence 84839, A
41	65	18.4	601	4	US-09-949-016-84840	Sequence 84840, A
42	65	18.4	321022	4	US-09-949-016-11852	Sequence 11852, A
43	65	18.4	321022	4	US-09-949-016-14166	Sequence 14166, A
c 44	64.5	18.3	40429	4	US-08-311-731A-125	Sequence 125, App
c 45	64.5	18.3	421491	4	US-09-949-016-12805	Sequence 12805, A

ALIGNMENTS

```
RESULT 1
US-08-978-741-16
; Sequence 16, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  1100 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-16

Alignment Scores:
Pred. No.:              1.19e-39        Length:         1100
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-9 (1-61) x US-08-978-741-16 (1-1100)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 His 61
              |||
Db        181 CAT 183

RESULT 2
US-09-333-729A-16
; Sequence 16, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 16
;    LENGTH: 1100
;    TYPE: DNA
;    ORGANISM: Homo Sapien
US-09-333-729A-16

Alignment Scores:
Pred. No.:              1.19e-39        Length:         1100
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-9 (1-61) x US-09-333-729A-16 (1-1100)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 His 61
              |||
Db        181 CAT 183

RESULT 3
US-08-978-741-4
; Sequence 4, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1300 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-4

Alignment Scores:
Pred. No.:              1.5e-39         Length:         1300
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-9 (1-61) x US-08-978-741-4 (1-1300)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy         61 His 61
              |||
Db        316 CAT 318
```

```
RESULT 4
US-09-333-729A-6
; Sequence 6, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 6
;   LENGTH: 1300
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid insert.
US-09-333-729A-6

Alignment Scores:
Pred. No.:              1.5e-39       Length:          1300
Score:                  353.00        Matches:         61
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
Query Match:            100.00%       Indels:          0
DB:                     3             Gaps:            0

US-09-774-954-9 (1-61) x US-09-333-729A-6 (1-1300)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 195

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 255

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 315

Qy         61 His 61
              |||
Db        316 CAT 318

RESULT 5
US-08-978-741-1
; Sequence 1, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1514 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-1

Alignment Scores:
Pred. No.:              1.84e-39      Length:          1514
Score:                  353.00        Matches:         61
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
Query Match:            100.00%       Indels:          0
DB:                     3             Gaps:            0

US-09-774-954-9 (1-61) x US-08-978-741-1 (1-1514)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 His 61
              |||
Db        181 CAT 183

RESULT 6
US-09-333-729A-2
; Sequence 2, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 2
;   LENGTH: 1514
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-2

Alignment Scores:
Pred. No.:              1.84e-39      Length:          1514
Score:                  353.00        Matches:         61
Percent Similarity:     100.00%       Conservative:    0
Best Local Similarity:  100.00%       Mismatches:      0
Query Match:            100.00%       Indels:          0
DB:                     3             Gaps:            0

US-09-774-954-9 (1-61) x US-09-333-729A-2 (1-1514)
```

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 His 61
              |||
Db        181 CAT 183

RESULT 7
US-08-978-741-5
; Sequence 5, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11284 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-978-741-5

Alignment Scores:
Pred. No.:              2.81e-38        Length:         11284
Score:                  353.00          Matches:        61
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-9 (1-61) x US-08-978-741-5 (1-11284)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy         61 His 61
              |||
Db       4416 CAT 4418

RESULT 8
US-09-270-767-14353
; Sequence 14353, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14353
;   LENGTH: 1320
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
US-09-270-767-14353

Alignment Scores:
Pred. No.:              2.18e-21        Length:         1320
Score:                  223.00          Matches:        38
Percent Similarity:     91.11%          Conservative:   3
Best Local Similarity:  84.44%          Mismatches:     4
Query Match:            63.17%          Indels:         0
DB:                     4               Gaps:           0

US-09-774-954-9 (1-61) x US-09-270-767-14353 (1-1320)

Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGlnAlaAsp 26
              ||||||   |||||||||   ||||||||||||||||||||||||||||||||||||
Db         75 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCGAC 134

Qy         27 HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProPro 46
              ||||||||||||||||||||||||||||   ||||||||||||||   :::|||||||
Db        135 CACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCG 194

Qy         47 TrpIleGluTyrGln 51
              |||:::|||||||:::
Db        195 TGGGTGGAGTATCGT 209

RESULT 9
US-08-978-741-7
; Sequence 7, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
```

```
;       FILING DATE:  26-No. 6100076-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/792498
;       FILING DATE:  31
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Svoboda, Craig G.
;       REGISTRATION NUMBER:  39,044
;       REFERENCE/DOCKET NUMBER:  P1041P1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650/225-1489
;       TELEFAX:  650/952-9881
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  5009 base pairs
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
US-08-978-741-7

Alignment Scores:
Pred. No.:              6.64e-20        Length:         5009
Score:                  218.00          Matches:        39
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            61.76%          Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-9 (1-61) x US-08-978-741-7 (1-5009)

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 61

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 118

RESULT 10
US-09-333-729A-8
; Sequence 8, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 8
;   LENGTH: 5009
;   TYPE: DNA
;   ORGANISM: Homo Sapien
US-09-333-729A-8

Alignment Scores:
Pred. No.:              6.64e-20        Length:         5009
Score:                  218.00          Matches:        39
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            61.76%          Indels:         0
DB:                     3               Gaps:           0

US-09-774-954-9 (1-61) x US-09-333-729A-8 (1-5009)

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 61

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 118

RESULT 11
US-09-949-016-12127/c
; Sequence 12127, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12127
;   LENGTH: 19227
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-12127

Alignment Scores:
Pred. No.:              1.99e-05        Length:         19227
Score:                  120.00          Matches:        19
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  95.00%          Mismatches:     0
Query Match:            33.99%          Indels:         0
DB:                     4               Gaps:           0

US-09-774-954-9 (1-61) x US-09-949-016-12127 (1-19227)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       1718 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGTAAG 1659

RESULT 12
US-09-949-016-16285/c
; Sequence 16285, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16285
;   LENGTH: 19228
;   TYPE: DNA
;   ORGANISM: Human
US-09-949-016-16285

Alignment Scores:
Pred. No.:              1.99e-05        Length:         19228
Score:                  120.00          Matches:        19
Percent Similarity:     100.00%         Conservative:   1
Best Local Similarity:  95.00%          Mismatches:     0
Query Match:            33.99%          Indels:         0
```

```
DB:                     4              Gaps:          0

US-09-774-954-9 (1-61) x US-09-949-016-16285 (1-19228)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       1718 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGTAAG 1659

RESULT 13
US-09-949-016-13358/c
; Sequence 13358, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13358
;    LENGTH: 260247
;    TYPE: DNA
;    ORGANISM: Human
US-09-949-016-13358

Alignment Scores:
Pred. No.:              1.12e+03       Length:        260247
Score:                  75.50          Matches:       19
Percent Similarity:     70.27%         Conservative:  7
Best Local Similarity:  51.35%         Mismatches:    8
Query Match:            21.39%         Indels:        3
DB:                     4              Gaps:          2

US-09-774-954-9 (1-61) x US-09-949-016-13358 (1-260247)

Qy          8 ProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGlnAlaAspHis 27
              |||:::    ||||||:::|||||     |||   |||||:::::::   :::
Db     138282 CCCTCACCAATTCTACTGTTTTGTCCC------AGGCGCCTAGGAAATGAGAGCCAATAT 138229

Qy         28 PheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaVal 44
              ||||||   |||   |||||||||   ||||||   :::|||:::|||||||
Db     138228 TTTTTATGAAGCCCTGCTTTTGCT---CTTCTTAAACAAACAATTGCAGTT 138181

RESULT 14
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;    LENGTH: 4403765
;    TYPE: DNA
;    ORGANISM: Mycobacterium tuberculosis
;    FEATURE:
;    OTHER INFORMATION: CDC 1551
;    OTHER INFORMATION: "n" bases at various positions throughout the sequence
;    OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.:              2.1e+05        Length:        4403765
Score:                  71.00          Matches:       18
Percent Similarity:     47.62%         Conservative:  2
Best Local Similarity:  42.86%         Mismatches:    10
Query Match:            20.11%         Indels:        12
DB:                     3              Gaps:          2

US-09-774-954-9 (1-61) x US-09-103-840A-2 (1-4403765)

Qy          5 SerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGln 24
              |||||||||   |||      :::|||   |||||||||   ||||||
Db     222566 AGTTGGGACCTTGCT------GTCCTGGTATGCCCTTGTCGCGGCCGTGACAACGGTTCC 222619

Qy         25 AlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaVal 44
              |||                                  |||:::|||||||||||||
Db     222620 GCTGGT-----------------------------GCGTTGGACCGCACACTGGCCGCG 222649

Qy         45 ProPro 46
              ||||||
Db     222650 CCGCCG 222655

RESULT 15
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;    LENGTH: 4411529
;    TYPE: DNA
;    ORGANISM: Mycobacterium tuberculosis
;    OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.:              2.1e+05        Length:        4411529
Score:                  71.00          Matches:       18
Percent Similarity:     47.62%         Conservative:  2
Best Local Similarity:  42.86%         Mismatches:    10
Query Match:            20.11%         Indels:        12
DB:                     3              Gaps:          2

US-09-774-954-9 (1-61) x US-09-103-840A-1 (1-4411529)

Qy          5 SerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGln 24
              |||||||||   |||      :::|||   |||||||||   ||||||
Db     222457 AGTTGGGACCTTGCT------GTCCTGGTATGCCCTTGTCGCGGCCGTGACAACGGTTCC 222510

Qy         25 AlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaVal 44
              |||                                  |||:::|||||||||||||
Db     222511 GCTGGT-----------------------------GCGTTGGACCGCACACTGGCCGCG 222540

Qy         45 ProPro 46
              ||||||
Db     222541 CCGCCG 222546


Search completed: October 26, 2005, 15:52:24
```

Job time : 1666.48 secs

GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 26, 2005, 03:15:37 ; Search time 170.85 Seconds
(without alignments)
2113.573 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool_h/US09774954/runat_25102005_105431_6380/app_query.fasta_1.917
-DB=N_Geneseq -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09774954_@CGN_1_1_703_@runat_25102005_105431_6380 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	353	100.0	1514	2	AAV65632	Aav65632 Human hea
2	353	100.0	5218	13	ACN40746	Acn40746 Tumour-as
3	353	100.0	5266	10	ADF81754	Adf81754 Leukaemia
4	353	100.0	11284	2	AAV65633	Aav65633 Plasmid c
5	241.5	68.4	5230	4	AAK51510	Aak51510 Human pol
6	229.5	65.0	4850	4	AAK52494	Aak52494 Human pol
7	223	63.2	1209	4	ABL13947	Abl13947 Drosophil
8	218	61.8	5009	2	AAV65634	Aav65634 First Eco
c 9	188	53.3	3264	4	ABL13946	Abl13946 Drosophil
c 10	188	53.3	3793	4	ABL13894	Abl13894 Drosophil
11	120	34.0	760	3	AAZ98210	Aaz98210 Human sig
12	94	26.6	97081	12	ADQ97980	Adq97980 Human can
13	90	25.5	591	4	AAI18766	Aai18766 Probe #86
14	90	25.5	591	4	AAI28271	Aai28271 Probe #18
15	90	25.5	591	4	ABA76618	Aba76618 Human foe
16	90	25.5	591	4	ABA63761	Aba63761 Human foe
17	90	25.5	591	4	AAI43879	Aai43879 Probe #12
18	90	25.5	591	4	AAI57317	Aai57317 Probe #26
19	90	25.5	591	4	ABA30952	Aba30952 Probe #94
20	90	25.5	591	4	ABA41126	Aba41126 Probe #19
21	90	25.5	591	4	AAK51260	Aak51260 Human bon
22	90	25.5	591	4	AAK37996	Aak37996 Human bon
23	90	25.5	591	4	AAK25254	Aak25254 Human bra
24	90	25.5	591	4	AAK12275	Aak12275 Human bra
25	90	25.5	591	4	ABS37615	Abs37615 Human liv
26	90	25.5	591	4	ABS50816	Abs50816 Human liv
27	90	25.5	591	6	ABS24798	Abs24798 Human gen
28	90	25.5	591	6	ABS11999	Abs11999 Human gen
29	90	25.5	591	12	ACH91620	Ach91620 Human gen
c 30	75.5	21.4	277616	13	ABD32602	Abd32602 Human can
31	73	20.7	3545	4	ABL09670	Abl09670 Drosophil
c 32	72	20.4	430	8	ABX45469	Abx45469 Bovine ES
c 33	72	20.4	432	8	ABX36491	Abx36491 Bovine ES
c 34	72	20.4	447	8	ABX41524	Abx41524 Bovine ES
35	71	20.1	1186	8	ACA38576	Aca38576 Prokaryot
36	71	20.1	1242	8	ACA40300	Aca40300 Prokaryot
37	71	20.1	110000	4	AAI99682_02	Continuation (3 of
38	71	20.1	110000	4	AAI99683_02	Continuation (3 of
39	70.5	20.0	467	9	ACH41843	Ach41843 Human foe
c 40	70.5	20.0	510	12	ACH76865	Ach76865 Human gen
c 41	70.5	20.0	5983	4	AAK84657	Aak84657 Human imm
c 42	70	19.8	326002	13	ABD32843	Abd32843 Human can
43	69.5	19.7	5025	6	ABN59850	Abn59850 Novel hum
44	68.5	19.4	1155	2	AAT01082	Aat01082 2-Alpha-f
45	68.5	19.4	1155	2	AAQ98461	Aaq98461 GDP-L-fuc

ALIGNMENTS

RESULT 1
AAV65632
ID AAV65632 standard; DNA; 1514 BP.
XX
AC AAV65632;
XX
DT 16-DEC-1998 (first entry)
XX
DE Human heart O-fucosyltransferase encoding DNA.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT misc_feature 1. .1100
FT /*tag= b
FT /note= "this actively expressed O-fucosyltransferase
FT sequence is claimed for in claim 9"
FT CDS 1. .1098
FT /*tag= a
FT /product= "human heart O-fucosyltransferase"
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.

XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
DR P-PSDB; AAW80571.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Claim 9; Fig 12A; 90pp; English.
XX
CC This DNA encodes a human heart O-fucosyltransferase that can glycosylate
CC an epidermal growth factor (EGF) domain of a polypeptide with an
CC activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.
CC mutants with increased affinity for the EFG domains, are used in
CC diagnosis and treatment of conditions associated with overexpression of O
CC -fucosyltransferase, to promote survival of sensory (retinal) neurons.
CC Probes based on EGF domain polypeptide are used to detect gene
CC amplification and expression. The expression can also be determined at
CC the protein level using antibodies specific for O-fucosyltransferase
XX
SQ Sequence 1514 BP; 320 A; 435 C; 411 G; 348 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.56e-37	Length:	1514
Score:	353.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	2	Gaps:	0

US-09-774-954-9 (1-61) x AAV65632 (1-1514)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 60

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 120

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 180

Qy         61 His 61
              |||
Db        181 CAT 183
```

RESULT 2
ACN40746
ID ACN40746 standard; cDNA; 5218 BP.
XX
AC ACN40746;
XX
DT 18-NOV-2004 (first entry)
XX
DE Tumour-associated antigenic target (TAT) cDNA DNA326813, SEQ ID NO:5725.
XX
KW Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW tumour; diagnosis; cell proliferative disorder; breast cancer;
KW colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW central nervous system cancer; bladder cancer; pancreatic cancer;
KW cervical cancer; melanoma; leukaemia; hybridisation probe;
KW chromosome identification; chromosome mapping; gene mapping;
KW gene therapy; cytostatic; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2004030615-A2.
XX
PD 15-APR-2004.
XX
PF 29-SEP-2003; 2003WO-US028547.
XX
PR 02-OCT-2002; 2002US-0414971P.
XX
PA (GETH) GENENTECH INC.
XX
PI Wu TD, Zhang Z, Zhou Y;
XX
DR WPI; 2004-347921/32.
DR P-PSDB; ABM82223.
XX
PT New tumor-associated antigenic target polypeptides and nucleic acids,
PT useful in preparing a medicament for treating or detecting a
PT proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT prostate cancer or tumor.
XX
PS Claim 1; SEQ ID NO 5725; 7273pp; English.
XX
CC The invention relates to human tumour-associated antigenic target (TAT)
CC polypeptides, and their related nucleic acids. The TAT polypeptides are
CC overexpressed in cancer tissues compared to normal tissues, and may thus
CC serve as effective targets for the diagnosis and treatment of cancer in
CC mammals. The invention also relates to nucleic acid and polypeptide
CC sequences at least 80% identical to the TAT nucleic acids and
CC polypeptides; expression vectors and host cells comprising a TAT nucleic
CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC TAT polypeptide; and methods and compositions for the treatment or
CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC antibodies, antagonists, binding molecules and compositions are useful
CC for diagnosing or treating a cell proliferative disorder associated with
CC increased TAT expression, particularly cancers such as breast cancer,
CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC cancer, pancreatic cancer, cervical cancer, cancers of the central
CC nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC used as hybridisation probes, in chromosome and gene mapping, in
CC chromosome identification and in gene therapy. The present sequence
CC represents a TAT nucleic acid of the invention
XX
SQ Sequence 5218 BP; 1255 A; 1274 C; 1334 G; 1355 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	2.22e-36	Length:	5218
Score:	353.00	Matches:	61
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	13	Gaps:	0

US-09-774-954-9 (1-61) x ACN40746 (1-5218)

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 178

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 238

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 298

Qy         61 His 61
              |||
Db        299 CAT 301
```

```
RESULT 3
ADF81754
ID   ADF81754 standard; DNA; 5266 BP.
XX
AC   ADF81754;
XX
DT   26-FEB-2004  (first entry)
XX
DE   Leukaemia-related DNA sequence #2310.
XX
KW   Cytostatic; Gene therapy; leukaemia; ss.
XX
OS   Unidentified.
XX
PN   WO2003039443-A2.
XX
PD   15-MAY-2003.
XX
PF   04-NOV-2002; 2002WO-EP012303.
XX
PR   05-NOV-2001; 2001EP-00126244.
PR   30-APR-2002; 2002EP-00009758.
XX
PA   (DEKR-) DEUT KREBSFORSCHUNGSZENTRUM.
PA   (UYLU-) UNIV LUDWIG MAXIMILIANS.
PA   (HAFE/) HAFERLACH T.
PA   (SCHO/) SCHOCH C.
PA   (KERN/) KERN W.
XX
PI   Haferlach T,  Schoch C,  Kern W,  Kohlmann A,  Schnittger S,  Dugas M;
PI   Eils R,  Brors B,  Mergenthaler S;
XX
DR   WPI; 2003-505037/47.
XX
PT   Determining the subtype of leukemia cells and whether a patient sample
PT   contains leukemia cells or other cells, useful for treating leukemia,
PT   comprises determining the expression profile of a group of markers in a
PT   patient sample.
XX
PS   Disclosure; SEQ ID NO 2310; 2938pp; English.
XX
CC   The present invention relates to a method (M1) for determining the
CC   subtype of leukaemia cells and whether a patient sample contains
CC   leukaemia cells. The method comprises determining the expression profile
CC   of a group of markers in a patient sample. The method is useful for
CC   determining the presence of leukaemia cells, its types or subtypes, and
CC   for the preparation of a medicament for treating leukaemia.
XX
SQ   Sequence 5266 BP; 1280 A; 1278 C; 1331 G; 1346 T; 0 U; 31 Other;

Alignment Scores:
Pred. No.:                2.25e-36        Length:          5266
Score:                    353.00          Matches:         61
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       10              Gaps:            0

US-09-774-954-9 (1-61) x ADF81754 (1-5266)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        132 ATGCCTGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 191

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        192 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 251

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        252 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 311

Qy         61 His 61
              |||
Db        312 CAT 314

RESULT 4
AAV65633
ID   AAV65633 standard; DNA; 11284 BP.
XX
AC   AAV65633;
XX
DT   16-DEC-1998  (first entry)
XX
DE   Plasmid construct for expression of human O-fucosyltransferase.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS   Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             4140. .5333
FT                   /*tag=  a
FT                   /note= "insert coding for human O-fucosyltransferase."
FT   misc_feature    4218. .4235
FT                   /*tag=  b
FT                   /note= "polyhistidine tag"
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
DR   P-PSDB; AAW80573.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example; Page 49-57; 90pp; English.
XX
CC   This represents the nucleotide sequence of the plasmid construct used for
CC   the expression of human heart O-fucosyltransferase. The human O-
CC   fucosyltransferase can glycosylate an epidermal growth factor (EGF)
CC   domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC   O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 11284 BP; 2947 A; 2602 C; 2624 G; 3111 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                5.95e-36        Length:          11284
Score:                    353.00          Matches:         61
Percent Similarity:       100.00%         Conservative:    0
Best Local Similarity:    100.00%         Mismatches:      0
Query Match:              100.00%         Indels:          0
DB:                       2               Gaps:            0

US-09-774-954-9 (1-61) x AAV65633 (1-11284)
```

```
Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4236 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGGCGC 4295

Qy         21 PheGlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArg 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4296 TTTGGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGT 4355

Qy         41 ThrLeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeu 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4356 ACCTTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTC 4415

Qy         61 His 61
              |||
Db       4416 CAT 4418
```

RESULT 5
AAK51510
ID AAK51510 standard; cDNA; 5230 BP.
XX
AC AAK51510;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 55.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM78377.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 1; Page 638-642; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC sequence listing were missing at the time of publication
XX
SQ Sequence 5230 BP; 1285 A; 1255 C; 1330 G; 1360 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1.87e-21	Length:	5230
Score:	241.50	Matches:	46
Percent Similarity:	81.67%	Conservative:	3
Best Local Similarity:	76.67%	Mismatches:	6
Query Match:	68.41%	Indels:	5
DB:	4	Gaps:	1

US-09-774-954-9 (1-61) x AAK51510 (1-5230)

```
Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet--------------GlyArgPhe 21
              |||  |||  :::  :::  |||  :::              |||||||||
Db        160 GACATGGCCACGTGGCCTACACATGCTCCAAGCGTGGAGGAAGGAGGCCAGGGGCGCTTT 219

Qy         22 GlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThr 41
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 GGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACC 279

Qy         42 LeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 TTGGCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 339
```

RESULT 6
AAK52494
ID AAK52494 standard; cDNA; 4850 BP.
XX
AC AAK52494;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2023.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US004098.
XX
PR 03-FEB-2000; 2000US-00496914.
PR 27-APR-2000; 2000US-00560875.
PR 20-JUN-2000; 2000US-00598075.
PR 19-JUL-2000; 2000US-00620325.
PR 01-SEP-2000; 2000US-00654936.
PR 15-SEP-2000; 2000US-00663561.
PR 20-OCT-2000; 2000US-00693325.
PR 30-NOV-2000; 2000US-00728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM79361.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT in diagnosis and gene therapy.
XX
PS Claim 1; Page 4424-4425; 6221pp; English.
XX

```
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 4850 BP; 1142 A; 1218 C; 1244 G; 1246 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                6.85e-20        Length:         4850
Score:                    229.50          Matches:        44
Percent Similarity:       80.00%          Conservative:   4
Best Local Similarity:    73.33%          Mismatches:     7
Query Match:              65.01%          Indels:         5
DB:                       4               Gaps:           1

US-09-774-954-9 (1-61) x AAK52494 (1-4850)

Qy           7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet--------------GlyArgPhe 21
               |||   |||   :::     :::   |||   :::                 |||||||||
Db         160 GACATGGCCACGTGGCCTACACATGCTCCAAGCGTGGAGGAAGGAGGCCAGGGGCGCTTT 219

Qy          22 GlyAsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThr 41
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db         220 GGGAACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTGAACCGTTCC 279

Qy          42 LeuAlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
               ||||||||||||   ||||||||||||||||||||||||||||||||||||||||||||
Db         280 TTGGCTGTCCCTTCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 339

RESULT 7
ABL13947
ID   ABL13947 standard; cDNA; 1209 BP.
XX
AC   ABL13947;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36323.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69844.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 36323; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1209 BP; 302 A; 313 C; 307 G; 287 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                8.58e-20        Length:         1209
Score:                    223.00          Matches:        38
Percent Similarity:       91.11%          Conservative:   3
Best Local Similarity:    84.44%          Mismatches:     4
Query Match:              63.17%          Indels:         0
DB:                       4               Gaps:           0

US-09-774-954-9 (1-61) x ABL13947 (1-1209)

Qy           7 AspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArgPheGlyAsnGlnAlaAsp 26
               ||||||   |||||||||   ||||||||||||||||||||||||||||||||||||||
Db          82 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGACGCTTTGGCAACCAGGCCGAC 141

Qy          27 HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProPro 46
               ||||||||||||||||||||||||||||   ||||||||||||||   :::||||||
Db         142 CACTTCCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCG 201

Qy          47 TrpIleGluTyrGln 51
               |||:::||||||:::
Db         202 TGGGTGGAGTATCGT 216

RESULT 8
AAV65634
ID   AAV65634 standard; cDNA; 5009 BP.
XX
AC   AAV65634;
XX
DT   16-DEC-1998  (first entry)
XX
DE   First EcoR1 nucleotide fragment of human KIAA0180.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart; ss.
XX
OS   Homo sapiens.
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
DR   P-PSDB; AAW80577.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example 1; Fig 11; 90pp; English.
```

```
XX
CC   This represents a first EcoR1 nucleotide fragment of human KIAA0180. This
CC   5009 basepairs partial cDNA encodes for a protein of unknown function
CC   from myeblast cel line KG-1. The invention provides a human heart O-
CC   fucosyltransferase that can glycosylate an epidermal growth factor (EGF)
CC   domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC   O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC   domains, are used in diagnosis and treatment of conditions associated
CC   with overexpression of O-fucosyltransferase, to promote survival of
CC   sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC   used to detect gene amplification and expression. The expression can also
CC   be determined at the protein level using antibodies specific for O-
CC   fucosyltransferase
XX
SQ   Sequence 5009 BP; 1235 A; 1195 C; 1268 G; 1311 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              2.47e-18        Length:         5009
Score:                  218.00          Matches:        39
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            61.76%          Indels:         0
DB:                     2               Gaps:           0

US-09-774-954-9 (1-61) x AAV65634 (1-5009)

Qy         23 AsnGlnAlaAspHisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 AACCAGGCCGATCACTTCTTGGGCTCTCTGGCATTTGCAAAGCTGCTAAACCGTACCTTG 61

Qy         43 AlaValProProTrpIleGluTyrGlnHisHisLysProProPheThrAsnLeuHis 61
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GCTGTCCCTCCTTGGATTGAGTACCAGCATCACAAGCCTCCTTTCACCAACCTCCAT 118

RESULT 9
ABL13946/c
ID   ABL13946 standard; cDNA; 3264 BP.
XX
AC   ABL13946;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36320.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69843.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 36320; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3264 BP; 904 A; 742 C; 758 G; 860 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.48e-14        Length:         3264
Score:                  188.00          Matches:        38
Percent Similarity:     64.06%          Conservative:   3
Best Local Similarity:  59.38%          Mismatches:     4
Query Match:            53.26%          Indels:         19
DB:                     4               Gaps:           1

US-09-774-954-9 (1-61) x ABL13946 (1-3264)

Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet---------------------- 18
              ||||||   |||||||||   ||||||||||||||
Db       2183 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGTAAGTTTCACTTGGAAGCAACC 2124

Qy         19 ------------------------------GlyArgPheGlyAsnGlnAlaAspHisPh 28
                                            ||||||||||||||||||||||||||||
Db       2123 GAAAAACGACAACTAAACCGCGCTTTTCAAAGGACGCTTTGGCAACCAGGCCGACCACTT 2064

Qy         28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProProTrpIl 48
              |||||||||||||||||||||||   ||||||||||||||   :::|||||||||::
Db       2063 CCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCGTGGGT 2004

Qy         48 eGluTyrGln 51
              :|||||:::
Db       2003 GGAGTATCGT 1994

RESULT 10
ABL13894/c
ID   ABL13894 standard; cDNA; 3793 BP.
XX
AC   ABL13894;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 36164.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB69791.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 36164; 21pp + Sequence Listing; English.
```

```
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3793 BP; 1020 A; 853 C; 950 G; 970 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                1.79e-14        Length:          3793
Score:                    188.00          Matches:         38
Percent Similarity:       64.06%          Conservative:    3
Best Local Similarity:    59.38%          Mismatches:      4
Query Match:              53.26%          Indels:          19
DB:                       4               Gaps:            1

US-09-774-954-9 (1-61) x ABL13894 (1-3793)

Qy          7 AspProAlaGlyTyrLeuLeuTyrCysProCysMet----------------------- 18
              ||||||   |||||||||   ||||||||||||||
Db        675 GATCCCAATGGCTACCTCACCTACTGTCCGTGTATGGGTAAGTTTCACTTGGAAGCAACC 616

Qy         19 -------------------------------GlyArgPheGlyAsnGlnAlaAspHisPh 28
                                             ||||||||||||||||||||||||||||
Db        615 GAAAAACGACAACTAAACCGCGCTTTTCAAAGGACGCTTTGGCAACCAGGCCGACCACTT 556

Qy         28 eLeuGlySerLeuAlaPheAlaLysLeuLeuAsnArgThrLeuAlaValProProTrpIl 48
              |||||||||||||||||||||||||   ||||||||||||||   :::||||||||::
Db        555 CCTGGGATCATTGGCCTTCGCCAAGGCGCTTAATCGCACCCTGATCCTGCCGCCGTGGGT 496

Qy         48 eGluTyrGln 51
              :||||||:::
Db        495 GGAGTATCGT 486

RESULT 11
AAZ98210
ID   AAZ98210 standard; cDNA; 760 BP.
XX
AC   AAZ98210;
XX
DT   11-MAY-2000  (first entry)
XX
DE   Human signal peptide containing protein HSPP-102 cDNA SEQ ID NO:236.
XX
KW   Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
KW   inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW   antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;
KW   antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW   reproductive disorder; developmental disorder; arteriosclerosis;
KW   cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW   asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW   Parkinson's disease; Huntington's diseases; ovulatory defect;
KW   muscular dystrophy; ss.
XX
OS   Homo sapiens.
XX
PN   WO200000610-A2.
XX
PD   06-JAN-2000.
XX
PF   25-JUN-1999;   99WO-US014484.
XX
PR   26-JUN-1998;   98US-0090762P.
PR   31-JUL-1998;   98US-0094983P.
PR   01-OCT-1998;   98US-0102686P.
PR   11-DEC-1998;   98US-0112129P.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Lal P,  Tang YT,  Gorgone GA,  Corley NC,  Guegler KJ,  Baughn MR;
PI   Akerblom IE,  Au-Young J,  Yue H,  Patterson C,  Reddy R,  Hillman JL;
PI   Bandman O;
XX
DR   WPI; 2000-160673/14.
DR   P-PSDB; AAY87325.
XX
PT   New human signal peptide-containing proteins useful in treatment,
PT   prevention and diagnosis of e.g. cancer, inflammation and cardiovascular
PT   disease.
XX
PS   Claim 9; Page 307; 327pp; English.
XX
CC   AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the
CC   human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
CC   anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,
CC   neuroprotective, cardiovascular and antiasthmatic activities, and can be
CC   used in gene therapy. HSPPs can be used to treat or prevent disorders
CC   associated with decreased activity or function of HSPP. Antagonists of
CC   HSPP are used to treat or prevent disorders associated with increased
CC   activity or function of HSPP. Such diseases include cell proliferation
CC   (including cancer), inflammation, cardiovascular, neurological,
CC   reproductive or developmental disorders, (e.g. arteriosclerosis,
CC   cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
CC   asthma, Crohn's disease, microbial or other infections, congestive or
CC   ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
CC   diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
CC   nucleic acids can be used for the recombinant production of HSPP, for
CC   detecting HSPP in standard hybridisation and amplification assays (for
CC   diagnosis and monitoring), in gene therapy, as antisense, triplex-forming
CC   or ribozyme therapeutics, for detecting related sequences or genetic
CC   variations, and for chromosomal mapping. HSPP are also used to raise
CC   specific antibodies (Ab) and to screen for agonists and antagonists
CC   (potential therapeutic agents). Ab are used to diagnose, or monitor, HSPP
CC   -related diseases (in usual immunoassays), as therapeutic antagonists, in
CC   competitive drug screens, and for purification of HSPP from natural
CC   sources
XX
SQ   Sequence 760 BP; 171 A; 206 C; 220 G; 163 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                2.9e-06         Length:          760
Score:                    120.00          Matches:         19
Percent Similarity:       100.00%         Conservative:    1
Best Local Similarity:    95.00%          Mismatches:      0
Query Match:              33.99%          Indels:          0
DB:                       3               Gaps:            0

US-09-774-954-9 (1-61) x AAZ98210 (1-760)

Qy          1 MetProAlaGlySerTrpAspProAlaGlyTyrLeuLeuTyrCysProCysMetGlyArg 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        111 ATGCCCGCGGGCTCCTGGGACCCGGCCGGTTACCTGCTCTACTGCCCCTGCATGGGTAAG 170

RESULT 12
ADQ97980
ID   ADQ97980 standard; DNA; 97081 BP.
XX
AC   ADQ97980;
XX
DT   07-OCT-2004  (first entry)
XX
DE   Human cancer associated sequence HD11-034, SEQ ID 957.
XX
KW   Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS   Homo sapiens.
XX
PN   WO2004060304-A2.
XX
PD   22-JUL-2004.
```

```
XX
PF   22-DEC-2003; 2003WO-US041389.
XX
PR   27-DEC-2002; 2002US-00330773.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-543781/52.
XX
PT   New isolated cancer associated nucleic acids comprising at least 10
PT   contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT   cancers such as leukemia and lymphoma.
XX
PS   Claim 1; SEQ ID NO 957; 199pp; English.
XX
CC   The present invention relates to cancer associated sequences (ADQ97025-
CC   ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC   treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC   data for this patent did not form part of the printed specification, but
CC   was obtained in electronic formate directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 97081 BP; 23547 A; 23101 C; 24653 G; 25780 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              4.33          Length:         97081
Score:                  94.00         Matches:        21
Percent Similarity:     43.84%        Conservative:   11
Best Local Similarity:  28.77%        Mismatches:     19
Query Match:            26.63%        Indels:         22
DB:                     12            Gaps:           2

US-09-774-954-9 (1-61) x ADQ97980 (1-97081)

Qy          2 ProAlaGlySerTrpAspProAlaGlyTyrLeu---LeuTyrCysProCysMetGlyArg 20
              |||   ||||||||||||||||||||||   |||   |||:::      |||
Db      66320 CCGCAGGGCAGTTGGGACCCCGCGGGCCTGCTCAGACTACACGGGTCGTGCCCAGCGCCA 66379

Qy         21 PheGlyAsnGlnAlaAsp----------------------------------------- 26
              ||||||:::   :::|||
Db      66380 TTTGGGCACTCCAGTGACCAGGGCCTCGGGTCCTGCACACCTCCCATGAGGAGAGATGTC 66439

Qy         27 --------------------HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsn 39
                               |||     ||||||:::   :::   ::::::
Db      66440 TTCCTCACATGGAATCAAAGCCACATCTGTGGTTCTATCCTCTGGGACCAGGTCTGCAGA 66499

Qy         40 ArgThrLeuAlaValProProTrpIleGluTyrGlnHis 52
                 |||      |||||||||:::   :::::::::
Db      66500 GCTGGTTTAGGTCATCCTCCTTGGGTACTCCACAGAAAT 66538

RESULT 13
AAI18766
ID   AAI18766 standard; DNA; 591 BP.
XX
AC   AAI18766;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #8699 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 25; SEQ ID NO 8699; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging of
CC   diseases of the cervix, notably cervical cancer. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 591 BP; 139 A; 145 C; 156 G; 151 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0.0217        Length:         591
Score:                  90.00         Matches:        21
Percent Similarity:     41.18%        Conservative:   7
Best Local Similarity:  30.88%        Mismatches:     18
Query Match:            25.50%        Indels:         22
DB:                     4             Gaps:           2

US-09-774-954-9 (1-61) x AAI18766 (1-591)

Qy          2 ProAlaGlySerTrpAspProAlaGlyTyrLeu---LeuTyrCysProCysMetGlyArg 20
              |||   ||||||||||||||||||||||   |||   |||:::      |||
Db        387 CCGCAGGGCAGTTGGGACCCCGCGGGCCTGCTCAGACTACACGGGTCGTGCCCAGCGCCA 446

Qy         21 PheGlyAsnGlnAlaAsp----------------------------------------- 26
              ||||||:::   :::|||
Db        447 TTTGGGCACTCCAGTGACCAGGGCCTCGGGTCCTGCACACCTCCCATGAGGAGAGATGTC 506

Qy         27 --------------------HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsn 39
                               |||     ||||||:::   :::   ::::::
Db        507 TTCCTCACATGGAATCAAAGCCACATCTGTGGTTCTATCCTCTGGGACCAGGTCTGCAGA 566

Qy         40 ArgThrLeuAlaValProProTrp 47
                 |||      |||||||||
Db        567 GCTGGTTTAGGTCATCCTCCTTGG 590

RESULT 14
AAI28271
ID   AAI28271 standard; DNA; 591 BP.
XX
AC   AAI28271;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #18204 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
```

```
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 25; SEQ ID NO 18204; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging of
CC   diseases of the cervix, notably cervical cancer. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 591 BP; 139 A; 145 C; 156 G; 151 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0.0217        Length:          591
Score:                  90.00         Matches:         21
Percent Similarity:     41.18%        Conservative:    7
Best Local Similarity:  30.88%        Mismatches:      18
Query Match:            25.50%        Indels:          22
DB:                     4             Gaps:            2

US-09-774-954-9 (1-61) x AAI28271 (1-591)

Qy          2 ProAlaGlySerTrpAspProAlaGlyTyrLeu---LeuTyrCysProCysMetGlyArg 20
              |||   ||||||||||||||||||||||   |||   |||:::      |||
Db        387 CCGCAGGGCAGTTGGGACCCCGCGGGCCTGCTCAGACTACACGGGTCGTGCCCAGCGCCA 446

Qy         21 PheGlyAsnGlnAlaAsp------------------------------------------ 26
              ||||||:::   :::|||
Db        447 TTTGGGCACTCCAGTGACCAGGGCCTCGGGTCCTGCACACCTCCCATGAGGAGAGATGTC 506

Qy         27 --------------------HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsn 39
                                 |||      ||||||:::   :::   ::::::
Db        507 TTCCTCACATGGAATCAAAGCCACATCTGTGGTTCTATCCTCTGGGACCAGGTCTGCAGA 566

Qy         40 ArgThrLeuAlaValProProTrp 47
                 |||      |||||||||
Db        567 GCTGGTTTAGGTCATCCTCCTTGG 590

RESULT 15
ABA76618
ID   ABA76618 standard; DNA; 591 BP.
XX
AC   ABA76618;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #24923.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 4; SEQ ID NO 24923; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a single exon nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 591 BP; 139 A; 145 C; 156 G; 151 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              0.0217        Length:          591
Score:                  90.00         Matches:         21
Percent Similarity:     41.18%        Conservative:    7
Best Local Similarity:  30.88%        Mismatches:      18
Query Match:            25.50%        Indels:          22
DB:                     4             Gaps:            2

US-09-774-954-9 (1-61) x ABA76618 (1-591)

Qy          2 ProAlaGlySerTrpAspProAlaGlyTyrLeu---LeuTyrCysProCysMetGlyArg 20
              |||   ||||||||||||||||||||||   |||   |||:::      |||
Db        387 CCGCAGGGCAGTTGGGACCCCGCGGGCCTGCTCAGACTACACGGGTCGTGCCCAGCGCCA 446

Qy         21 PheGlyAsnGlnAlaAsp------------------------------------------ 26
              ||||||:::   :::|||
Db        447 TTTGGGCACTCCAGTGACCAGGGCCTCGGGTCCTGCACACCTCCCATGAGGAGAGATGTC 506

Qy         27 --------------------HisPheLeuGlySerLeuAlaPheAlaLysLeuLeuAsn 39
                                 |||      ||||||:::   :::   ::::::
Db        507 TTCCTCACATGGAATCAAAGCCACATCTGTGGTTCTATCCTCTGGGACCAGGTCTGCAGA 566

Qy         40 ArgThrLeuAlaValProProTrp 47
                 |||      |||||||||
Db        567 GCTGGTTTAGGTCATCCTCCTTGG 590


Search completed: October 26, 2005, 09:52:01
Job time : 190.85 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:21:43 ; Search time 7.64066 Seconds
(without alignments)
768.157 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	123	34.8	474	2	T15511	hypothetical prote
2	78	22.1	354	2	I51266	opsin, green-sensi
3	67	19.0	355	2	A42347	opsin, green-sensi
4	66	18.7	355	2	I51319	RH2 opsin - green
5	64.5	18.3	219	2	S72721	amidotransferase h
6	64.5	18.3	223	2	B86968	conserved hypothet
7	64	18.1	212	2	D82431	CbbY family protei
8	62.5	17.7	603	2	A87342	conserved hypothet
9	62	17.6	349	2	B45229	opsin, green-sensi
10	62	17.6	353	1	JN0120	rhodopsin - Japane
11	61.5	17.4	349	2	A45229	opsin, green-sensi
12	60.5	17.1	767	2	D95101	SpoE family protei
13	60.5	17.1	767	2	E97969	cell division prot
14	59.5	16.9	198	2	C70570	hypothetical prote
15	59.5	16.9	492	2	T41004	hypothetical prote
16	59.5	16.9	569	2	E86957	probable cholester
17	59.5	16.9	585	2	S72824	cholesterol oxidas
18	59	16.7	2482	2	I48922	cation-independent
19	59	16.7	2483	1	A49617	insulin-like growt
20	58.5	16.6	430	2	AB2316	hypothetical prote
21	58.5	16.6	504	2	B40829	activin receptor i
22	58.5	16.6	528	2	C40829	activin receptor i
23	58	16.4	225	2	B88712	protein C17H12.10
24	58	16.4	365	2	A36047	galactoside 2-alph
25	58	16.4	386	2	D69040	conserved hypothet
26	58	16.4	478	2	G71443	hypothetical prote
27	58	16.4	912	2	F71433	probable growth re
28	57.5	16.3	461	2	T00918	hypothetical prote
29	57.5	16.3	491	2	S76943	hypothetical prote
30	57.5	16.3	1329	2	T33136	hypothetical prote
31	57	16.1	199	2	AG0707	conserved hypothet
32	57	16.1	352	1	S40688	opsin, rod - Pomat
33	57	16.1	354	1	S27231	rhodopsin - northe
34	57	16.1	562	2	T40967	hypothetical prote
35	56.5	16.0	254	2	H83353	probable permease
36	56.5	16.0	284	2	AB1919	hypothetical prote
37	56.5	16.0	513	2	G84770	hypothetical prote
38	56	15.9	307	2	AF2402	hypothetical prote
39	56	15.9	328	2	A70127	phosphate ABC tran
40	56	15.9	351	2	S29152	rhodopsin - chicke
41	56	15.9	355	1	A46191	iodopsin homolog -
42	56	15.9	421	2	I49375	serotonin receptor
43	56	15.9	422	2	I38209	serotonin receptor
44	56	15.9	422	2	JH0315	serotonin receptor
45	56	15.9	588	2	E64715	penicillin-binding

ALIGNMENTS

RESULT 1
T15511
hypothetical protein C15C7.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15511
R;Leimbach, D.
submitted to the EMBL Data Library, November 1995
A;Description: The sequence of C. elegans cosmid C15C7.
A;Reference number: Z18363
A;Accession: T15511
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-474 <LEI>
A;Cross-references: EMBL:U41528; NID:g1109795; PID:g1109800; PIDN:AAA83156.1; CESP:C15C7.
C;Genetics:
A;Gene: CESP:C15C7.1
A;Introns: 31/3; 67/2; 106/2; 166/1; 202/3; 235/3; 364/3; 410/3

Query Match 34.8%; Score 123; DB 2; Length 474;
Best Local Similarity 67.6%; Pred. No. 1.3e-07;
Matches 23; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

```
Qy         19 GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQH 52
              |||||| | ||| ||||| |:||| :| :||::|
Db        166 GRFGNQVDQFLGVLAFAKALDRTLVLPNFIEFKH 199
```

RESULT 2
I51266
opsin, green-sensitive - Mexican tetra
C;Species: Astyanax mexicanus (Mexican tetra)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: I51266
R;Register, E.A.; Yokoyama, R.; Yokoyama, S.
J. Mol. Evol. 39, 268-273, 1994
A;Title: Multiple origins of the green-sensitive opsin genes in fish.
A;Reference number: I51266; MUID:95018302; PMID:7932788
A;Accession: I51266
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-354 <REG>
A;Cross-references: UNIPROT:P51474; GB:S75255; NID:g807171; PIDN:AAB32221.1; PID:g807172
C;Genetics:
A;Gene: rh1Af
A;Introns: 124/1; 180/2; 235/3; 315/3
C;Superfamily: vertebrate rhodopsin
C;Keywords: chromoprotein; eye; G protein-coupled receptor; lipoprotein; phosphoprotein;
F;299/Binding site: retinal (Lys) (covalent) #status predicted

Query Match 22.1%; Score 78; DB 2; Length 354;
Best Local Similarity 33.3%; Pred. No. 0.046;

```
  Matches   15;  Conservative    7;  Mismatches   23;  Indels    0;  Gaps    0;

Qy         11 YLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKP 55
              |:: |  || |   | | || : |   :  | | || : :  : |
Db        139 YIVVCKPMGSFKFSASHALGGIGFTWFMAMTCAAPPLVGWSRYIP 183


RESULT 3
A42347
opsin, green-sensitive - chicken
N;Alternate names: green visual pigment
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: A42347; A46137
R;Wang, S.Z.; Adler, R.; Nathans, J.
Biochemistry 31, 3309-3315, 1992
A;Title: A visual pigment from chicken that resembles rhodopsin: amino acid sequence, ge
A;Reference number: A42347; MUID:92207951; PMID:1554715
A;Accession: A42347
A;Molecule type: DNA
A;Residues: 1-355 <WAN>
A;Cross-references: UNIPROT:P28683; GB:M88178; GB:J05369; NID:g212601; PIDN:AAA49036.1;
A;Note: sequence extracted from NCBI backbone (NCBIN:93619, NCBIN:93621, NCBIN:93623, NC
R;Okano, T.; Kojima, D.; Fukada, Y.; Shichida, Y.; Yoshizawa, T.
Proc. Natl. Acad. Sci. U.S.A. 89, 5932-5936, 1992
A;Title: Primary structures of chicken cone visual pigments: vertebrate rhodopsins have
A;Reference number: A46137; MUID:92335211; PMID:1385866
A;Accession: A46137
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-355 <OKA>
A;Cross-references: GB:M92038; NID:g211829; PIDN:AAA48786.1; PID:g211830
A;Experimental source: retina
A;Note: sequence extracted from NCBI backbone (NCBIP:108554)
C;Superfamily: vertebrate rhodopsin
C;Keywords: chromoprotein; eye; G protein-coupled receptor; glycoprotein; lipoprotein; p
F;37-61/Domain: transmembrane #status predicted <TM1>
F;74-96/Domain: transmembrane #status predicted <TM2>
F;114-133/Domain: transmembrane #status predicted <TM3>
F;153-175/Domain: transmembrane #status predicted <TM4>
F;203-230/Domain: transmembrane #status predicted <TM5>
F;253-276/Domain: transmembrane #status predicted <TM6>
F;285-309/Domain: transmembrane #status predicted <TM7>
F;296/Binding site: retinal (Lys) (covalent) #status predicted

  Query Match             19.0%;  Score 67;  DB 2;  Length 355;
  Best Local Similarity   26.2%;  Pred. No. 1.1;
  Matches   17;  Conservative    9;  Mismatches   23;  Indels   16;  Gaps    2;

Qy         11 YLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPP---WI-------------EYQHHK 54
              |:: |  || |   | | :  :||  ::  : | ||   |              :|  |
Db        136 YIVVCKPMGNFRFSATHAMMGIAFTWVMAFSCAAPPLFGWSRYMPEGMQCSCGPDYYTHN 195

Qy         55 PPFTN 59
              | : |
Db        196 PDYHN 200


RESULT 4
I51319
RH2 opsin - green anole
C;Species: Anolis carolinensis (green anole)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C;Accession: I51319
R;Kawamura, S.; Yokoyama, S.
J. Mol. Evol. 40, 594-600, 1995
A;Title: Paralogous origin of the rhodopsinlike opsin genes in lizards.
A;Reference number: I51319; MUID:95371134; PMID:7643409
A;Accession: I51319
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-355 <KAW>
A;Cross-references: UNIPROT:P51471; GB:S79167; NID:g1042072; PIDN:AAB35062.1; PID:g1042
C;Genetics:
A;Gene: rh2Ac
A;Introns: 121/1; 177/2; 232/3; 312/3
C;Superfamily: vertebrate rhodopsin
C;Keywords: chromoprotein; eye; G protein-coupled receptor; lipoprotein; phosphoprotein;
F;296/Binding site: retinal (Lys) (covalent) #status predicted
F;322,323/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             18.7%;  Score 66;  DB 2;  Length 355;
  Best Local Similarity   28.9%;  Pred. No. 1.5;
  Matches   13;  Conservative   10;  Mismatches   22;  Indels    0;  Gaps    0;

Qy         11 YLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKP 55
              |:: |  || |   | | |  ::|   :: : | || : :  : |
Db        136 YIVVCKPMGNFRFSATHALMGISFTWFMSFSCAAPPLLGWSRYIP 180


RESULT 5
S72721
amidotransferase hisH homolog - Mycobacterium leprae
N;Alternate names: B1177_C1_149
C;Species: Mycobacterium leprae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
C;Accession: S72721
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A;Description: Mycobacterium leprae cosmid B1177.
A;Reference number: S72694
A;Accession: S72721
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-219 <SMI>
A;Cross-references: UNIPROT:Q49637; EMBL:U00011; NID:g466807; PIDN:AAA17085.1; PID:g4668
C;Genetics:
A;Gene: hisH
C;Superfamily: conserved hypothetical protein HI1648

  Query Match             18.3%;  Score 64.5;  DB 2;  Length 219;
  Best Local Similarity   43.8%;  Pred. No. 1.4;
  Matches   14;  Conservative    4;  Mismatches   11;  Indels    3;  Gaps    1;

Qy         21 FGNQADHFLGSLAFAKLLNRTLAV---PPWIE 49
              || | | | | : || |::   ||    ||:|
Db        137 FGRQVDSFEGDIGFAGLVDPVRAVFIRAPWVE 168


RESULT 6
B86968
conserved hypothetical protein ML0474 [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C;Accession: B86968
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: B86968
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-223 <STO>
A;Cross-references: UNIPROT:Q9CCT5; GB:AL450380; NID:g13092704; PIDN:CAC29982.1; GSPDB:GI
C;Genetics:
A;Gene: ML0474
C;Superfamily: conserved hypothetical protein HI1648

  Query Match             18.3%;  Score 64.5;  DB 2;  Length 223;
  Best Local Similarity   43.8%;  Pred. No. 1.4;
  Matches   14;  Conservative    4;  Mismatches   11;  Indels    3;  Gaps    1;

Qy         21 FGNQADHFLGSLAFAKLLNRTLAV---PPWIE 49
```

```
                  || | | | | : || |::   ||    ||:|
Db        141 FGRQVDSFEGDIGFAGLVDPVRAVFIRAPWVE 172


RESULT 7
D82431
CbbY family protein VCA0662 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: D82431
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: D82431
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-212 <HEI>
A;Cross-references: UNIPROT:Q9KLS9; GB:AE004396; GB:AE003853; NID:g9658077; PIDN:AAF9656
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VCA0662
A;Map position: 2

  Query Match             18.1%;  Score 64;  DB 2;  Length 212;
  Best Local Similarity   35.2%;  Pred. No. 1.5;
  Matches   19;  Conservative    8;  Mismatches   21;  Indels    6;  Gaps    1;

Qy          8 PAGYLLYCPCMGR------FGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKP 55
              || | | |   |:       |:| |  |  |: |::|::  ||    : | |||
Db        105 PATYELLCQWQGKKKMAIGTGSQRDSALRLLSNAQVLDKFDAVVTASDVQQHKP 158


RESULT 8
A87342
conserved hypothetical protein CC0748 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: A87342
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: A87342
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-603 <STO>
A;Cross-references: GB:AE005673; NID:g13421981; PIDN:AAK22733.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0748

  Query Match             17.7%;  Score 62.5;  DB 2;  Length 603;
  Best Local Similarity   32.1%;  Pred. No. 7.5;
  Matches   17;  Conservative    5;  Mismatches   28;  Indels    3;  Gaps    1;

Qy          9 AGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              | :|  |    ||  : |  :| | |   ||    :  |   :  |||   ||
Db         50 AVFLAGCGLPERFAERRDFVVGELGFGSGLNIAALLDLW---RREKPPGGRLH 99


RESULT 9
B45229
opsin, green-sensitive (clone GFgr-2) - goldfish
C;Species: Carassius auratus (goldfish)
C;Date: 30-Apr-1993 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C;Accession: B45229
R;Johnson, R.L.; Grant, K.B.; Zankel, T.C.; Boehm, M.F.; Merbs, S.L.; Nathans, J.; Nakan
Biochemistry 32, 208-214, 1993
A;Title: Cloning and expression of goldfish opsin sequences.
A;Reference number: A45229; MUID:93120096; PMID:8418840
A;Accession: B45229
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-349 <JOH>
A;Cross-references: UNIPROT:P32312; GB:L11866; NID:g212959; PIDN:AAA49169.1; PID:g212960
A;Experimental source: retina
A;Note: sequence extracted from NCBI backbone (NCBIP:122266)
C;Superfamily: vertebrate rhodopsin
C;Keywords: chromoprotein; eye; G protein-coupled receptor; glycoprotein; lipoprotein; ph
F;2,15/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;279/Binding site: retinal (Lys) (covalent) #status predicted

  Query Match             17.6%;  Score 62;  DB 2;  Length 349;
  Best Local Similarity   26.7%;  Pred. No. 4.8;
  Matches   12;  Conservative    9;  Mismatches   24;  Indels    0;  Gaps    0;

Qy         11 YLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKP 55
              |:: |  || |   : |    :||  ::    | || : :  : |
Db        136 YIVVCKPMGSFKFSSTHASAGIAFTWVMAMACAAPPLVGWSRYIP 180


RESULT 10
JN0120
rhodopsin - Japanese lamprey
C;Species: Lampetra japonica (Japanese lamprey)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 09-Jul-2004
C;Accession: JN0120
R;Hisatomi, O.; Iwasa, T.; Tokunaga, F.; Yasui, A.
Biochem. Biophys. Res. Commun. 174, 1125-1132, 1991
A;Title: Isolation and characterization of lamprey rhodopsin cDNA.
A;Reference number: JN0120; MUID:91144585; PMID:1840482
A;Accession: JN0120
A;Molecule type: mRNA
A;Residues: 1-353 <HIS>
A;Cross-references: UNIPROT:P22671; GB:M63632; NID:g213348; PIDN:AAA49342.1; PID:g213349
A;Experimental source: liver
A;Note: the authors translated the codon AAC for residue 195 as Glu
C;Superfamily: vertebrate rhodopsin
C;Keywords: chromoprotein; eye; G protein-coupled receptor; glycoprotein; lipoprotein; ph
F;37-61/Domain: transmembrane #status predicted <TM1>
F;74-96/Domain: transmembrane #status predicted <TM2>
F;115-140/Domain: transmembrane #status predicted <TM3>
F;153-173/Domain: transmembrane #status predicted <TM4>
F;203-230/Domain: transmembrane #status predicted <TM5>
F;253-276/Domain: transmembrane #status predicted <TM6>
F;286-309/Domain: transmembrane #status predicted <TM7>
F;2,15/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;296/Binding site: retinal (Lys) (covalent) #status predicted
F;322,323/Binding site: palmitate (Cys) (covalent) #status predicted

  Query Match             17.6%;  Score 62;  DB 1;  Length 353;
  Best Local Similarity   31.9%;  Pred. No. 4.8;
  Matches   15;  Conservative    9;  Mismatches   19;  Indels    4;  Gaps    2;

Qy         11 YLLYCPCMG--RFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKP 55
              |:: |  ||  ||||   | :  :||  ::    | || : :  : |
Db        136 YIVICKPMGNFRFGN--THAIMGVAFTWIMALACAAPPLVGWSRYIP 180


RESULT 11
A45229
opsin, green-sensitive (clone GFgr-1) - goldfish
C;Species: Carassius auratus (goldfish)
C;Date: 30-Apr-1993 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C;Accession: A45229
R;Johnson, R.L.; Grant, K.B.; Zankel, T.C.; Boehm, M.F.; Merbs, S.L.; Nathans, J.; Nakani
Biochemistry 32, 208-214, 1993
A;Title: Cloning and expression of goldfish opsin sequences.
A;Reference number: A45229; MUID:93120096; PMID:8418840
A;Accession: A45229
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
```

```
A;Residues: 1-349 <JOH>
A;Cross-references: UNIPROT:P32311; GB:L11865; NID:g212957; PIDN:AAA49168.1; PID:g212958
A;Experimental source: retina
A;Note: sequence extracted from NCBI backbone (NCBIP:122265)
C;Superfamily: vertebrate rhodopsin
C;Keywords: chromoprotein; eye; G protein-coupled receptor; glycoprotein; lipoprotein; p
F;2,15/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;296/Binding site: retinal (Lys) (covalent) #status predicted

  Query Match             17.4%;  Score 61.5;  DB 2;  Length 349;
  Best Local Similarity   30.2%;  Pred. No. 5.5;
  Matches   13;  Conservative    6;  Mismatches   21;  Indels    3;  Gaps    1;

Qy         11 YLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPP---WIEY 50
              |:: |  || |   : |    :||  ::    | ||   |  |
Db        136 YIVVCKPMGSFKFSSSHAFAGIAFTWVMALACAAPPLFGWSRY 178


RESULT 12
D95101
SpoE family protein [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C;Accession: D95101
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: D95101
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-767 <KUR>
A;Cross-references: UNIPROT:Q8DQ94; GB:AE005672; PIDN:AAK75005.1; PID:g14972351; GSPDB:G
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP0878
C;Superfamily: Bacillus subtilis DNA translocase spoIIIE

  Query Match             17.1%;  Score 60.5;  DB 2;  Length 767;
  Best Local Similarity   36.6%;  Pred. No. 17;
  Matches   15;  Conservative    5;  Mismatches   18;  Indels    3;  Gaps    1;

Qy         10 GYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEY 50
              |  || |    | |   :|:||:    :|  :| | ||  |
Db        142 GVALYIPTAFLFSNIGTYFIGSIL---ILVGSLLVSPWSVY 179


RESULT 13
E97969
cell division protein [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C;Accession: E97969
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: E97969
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-767 <KUR>
A;Cross-references: UNIPROT:Q8DQ94; GB:AE007317; PIDN:AAK99585.1; PID:g15458378; GSPDB:G
C;Genetics:
A;Gene: ftsK
C;Superfamily: Bacillus subtilis DNA translocase spoIIIE

  Query Match             17.1%;  Score 60.5;  DB 2;  Length 767;
  Best Local Similarity   36.6%;  Pred. No. 17;
  Matches   15;  Conservative    5;  Mismatches   18;  Indels    3;  Gaps    1;

Qy         10 GYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEY 50
              |  || |    | |   :|:||:    :|  :| | ||  |
Db        142 GVALYIPTAFLFSNIGTYFIGSIL---ILVGSLLVSPWSVY 179


RESULT 14
C70570
hypothetical protein Rv2604c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C;Accession: C70570
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome s
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: C70570
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-198 <COL>
A;Cross-references: UNIPROT:O06210; GB:Z95387; GB:AL123456; NID:g3261763; PIDN:CAB08636.1
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv2604c
C;Superfamily: conserved hypothetical protein HI1648

  Query Match             16.9%;  Score 59.5;  DB 2;  Length 198;
  Best Local Similarity   43.8%;  Pred. No. 5.3;
  Matches   14;  Conservative    4;  Mismatches   11;  Indels    3;  Gaps    1;

Qy         21 FGNQADHFLGSLAFAKLLNRTLAV---PPWIE 49
              ||:| | | | : || | :    ||    ||:|
Db        116 FGSQVDSFEGDIEFAGLDDPVRAVFIRAPWVE 147


RESULT 15
T41004
hypothetical protein SPCC1494.03 - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T41004
R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, June 1998
A;Reference number: Z21963
A;Accession: T41004
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-492 <WOO>
A;Cross-references: UNIPROT:O60077; EMBL:AL023776; PIDN:CAA19301.1; GSPDB:GN00067; SPDB:S
A;Experimental source: strain 972h-; cosmid c1494
C;Genetics:
A;Gene: SPDB:SPCC1494.03
A;Map position: 2

  Query Match             16.9%;  Score 59.5;  DB 2;  Length 492;
  Best Local Similarity   41.2%;  Pred. No. 14;
  Matches   14;  Conservative    8;  Mismatches    9;  Indels    3;  Gaps    2;

Qy         28 FLGSLAFAKLLNRTLAVPPWIEYQHH-KPPFTNL 60
              :  :| |::||  ||:  ||::||  | || |:
Db        106 YRAQVADAQILN--LAITYWIDWQHKLKAPFFNM 137


Search completed: October 25, 2005, 15:36:44
Job time : 8.64066 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:22:39 ; Search time 35.8234 Seconds
(without alignments)
710.936 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1862994 seqs, 417510619 residues

Total number of hits satisfying chosen parameters: 1862994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	353	100.0	61	11	US-09-774-954-9	Sequence 9, Appli
2	353	100.0	365	11	US-09-774-954-2	Sequence 2, Appli
3	353	100.0	388	14	US-10-301-822-169	Sequence 169, App
4	353	100.0	397	11	US-09-774-954-6	Sequence 6, Appli
5	295	83.6	61	11	US-09-774-954-3	Sequence 3, Appli
6	223	63.2	402	20	US-11-097-143-18162	Sequence 18162, A
7	218	61.8	343	11	US-09-774-954-17	Sequence 17, Appl
8	123	34.8	474	11	US-09-774-954-8	Sequence 8, Appli
9	120	34.0	150	18	US-10-820-474A-102	Sequence 102, App
10	101	28.6	28	11	US-09-774-954-14	Sequence 14, Appl
11	86	24.4	110	16	US-10-425-115-327031	Sequence 327031,
12	80.5	22.8	152	15	US-10-424-599-266118	Sequence 266118,
13	77.5	22.0	189	16	US-10-425-115-201329	Sequence 201329,
14	77.5	22.0	538	16	US-10-437-963-113919	Sequence 113919,
15	73.5	20.8	601	14	US-10-156-761-12551	Sequence 12551, A
16	64.5	18.3	223	15	US-10-282-122A-63705	Sequence 63705, A
17	63.5	18.0	191	15	US-10-282-122A-62907	Sequence 62907, A
18	63.5	18.0	580	15	US-10-369-493-8280	Sequence 8280, Ap
19	63	17.8	156	16	US-10-425-115-282566	Sequence 282566,
20	62.5	17.7	1271	16	US-10-437-963-180066	Sequence 180066,
21	62	17.6	412	16	US-10-856-499-2258	Sequence 2258, Ap
22	61.5	17.4	119	16	US-10-437-963-148755	Sequence 148755,
23	61.5	17.4	543	15	US-10-424-599-145078	Sequence 145078,
24	61.5	17.4	1262	16	US-10-437-963-117000	Sequence 117000,
25	61	17.3	75	16	US-10-437-963-198684	Sequence 198684,
26	61	17.3	117	16	US-10-425-115-235644	Sequence 235644,
27	61	17.3	146	15	US-10-424-599-233161	Sequence 233161,
28	60.5	17.1	741	17	US-10-472-928-1698	Sequence 1698, Ap
29	60.5	17.1	763	18	US-10-617-320-3610	Sequence 3610, Ap
30	60.5	17.1	767	9	US-09-815-242-13343	Sequence 13343, A
31	60.5	17.1	767	9	US-09-815-242-13598	Sequence 13598, A
32	60.5	17.1	767	10	US-09-769-787-96	Sequence 96, Appl
33	60.5	17.1	767	15	US-10-282-122A-73911	Sequence 73911, A
34	60.5	17.1	783	9	US-09-775-978-2	Sequence 2, Appli
35	60	17.0	545	15	US-10-424-599-174345	Sequence 174345,
36	60	17.0	557	15	US-10-425-114-49507	Sequence 49507, A
37	59.5	16.9	169	15	US-10-424-599-195444	Sequence 195444,
38	59.5	16.9	198	15	US-10-282-122A-61790	Sequence 61790, A
39	59.5	16.9	198	15	US-10-282-122A-62362	Sequence 62362, A
40	59.5	16.9	198	15	US-10-282-122A-64747	Sequence 64747, A
41	59.5	16.9	398	9	US-09-796-338A-17	Sequence 17, Appl
42	59.5	16.9	398	14	US-10-282-837-17	Sequence 17, Appl
43	59.5	16.9	398	14	US-10-225-567A-477	Sequence 477, App
44	59.5	16.9	398	14	US-10-145-586-17	Sequence 17, Appl
45	59.5	16.9	398	14	US-10-241-220-104	Sequence 104, App

ALIGNMENTS

```
RESULT 1
US-09-774-954-9
; Sequence 9, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
```

```
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 9:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 61 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-774-954-9

  Query Match             100.0%;  Score 353;  DB 11;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-37;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 2
US-09-774-954-2
; Sequence 2, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 2:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 365 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-774-954-2

  Query Match             100.0%;  Score 353;  DB 11;  Length 365;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-36;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 3
US-10-301-822-169
; Sequence 169, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
;  APPLICANT:  Burgart, Lawrence J.
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
;  CURRENT APPLICATION NUMBER: US/10/301,822
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 169
;   LENGTH: 388
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-10-301-822-169

  Query Match             100.0%;  Score 353;  DB 14;  Length 388;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-36;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 83

Qy         61 H 61
              |
Db         84 H 84


RESULT 4
US-09-774-954-6
; Sequence 6, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
```

```
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/774,954
;            FILING DATE: 30-Jan-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/978,741
;            FILING DATE: 26-NOV-1997
;            APPLICATION NUMBER: 08/792,498
;            FILING DATE: 31-JAN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 6:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 397 amino acids
;            TYPE: Amino Acid
;            TOPOLOGY: Linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-774-954-6

  Query Match             100.0%;  Score 353;  DB 11;  Length 397;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-36;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         33 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 92

Qy         61 H 61
              |
Db         93 H 93


RESULT 5
US-09-774-954-3
; Sequence 3, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;       APPLICANT: Yang Wang, Michael W. Spellman
;       TITLE OF INVENTION: O-Fucosyltransferase
;       NUMBER OF SEQUENCES: 17
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genentech, Inc.
;            STREET: 1 DNA Way
;            CITY: South San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94080
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: WinPatin (Genentech)
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/774,954
;            FILING DATE: 30-Jan-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/978,741
;            FILING DATE: 26-NOV-1997
;            APPLICATION NUMBER: 08/792,498
;            FILING DATE: 31-JAN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;   INFORMATION FOR SEQ ID NO: 3:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 61 amino acids
;            TYPE: Amino Acid
;            TOPOLOGY: Linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-774-954-3

  Query Match             83.6%;  Score 295;  DB 11;  Length 61;
  Best Local Similarity   91.5%;  Pred. No. 6e-30;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||
Db          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 6
US-11-097-143-18162
; Sequence 18162, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18162
;   LENGTH: 402
;   TYPE: PRT
;   ORGANISM: DROSOPHILA
US-11-097-143-18162

  Query Match             63.2%;  Score 223;  DB 20;  Length 402;
  Best Local Similarity   84.4%;  Pred. No. 7.9e-20;
  Matches   38;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          7 DPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQ 51
              || ||| |||||||||||||||||||||| ||||| :|||:||:
Db         28 DPNGYLTYCPCMGRFGNQADHFLGSLAFAKALNRTLILPPWVEYR 72


RESULT 7
US-09-774-954-17
; Sequence 17, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;       APPLICANT: Yang Wang, Michael W. Spellman
;       TITLE OF INVENTION: O-Fucosyltransferase
;       NUMBER OF SEQUENCES: 17
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genentech, Inc.
```

```
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;    INFORMATION FOR SEQ ID NO: 17:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 343 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-774-954-17

  Query Match             61.8%;  Score 218;  DB 11;  Length 343;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-19;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         23 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              |||||||||||||||||||||||||||||||||||||||
Db          1 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 39


RESULT 8
US-09-774-954-8
; Sequence 8, Application US/09774954
; Publication No. US20040241645A1
;   GENERAL INFORMATION:
;        APPLICANT: Yang Wang, Michael W. Spellman
;        TITLE OF INVENTION: O-Fucosyltransferase
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 1 DNA Way
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: WinPatin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/774,954
;             FILING DATE: 30-Jan-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/978,741
;             FILING DATE: 26-NOV-1997
;             APPLICATION NUMBER: 08/792,498
;             FILING DATE: 31-JAN-1997
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Svoboda, Craig G.
;             REGISTRATION NUMBER: 39,044
;             REFERENCE/DOCKET NUMBER: P1041P1
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 650/225-1489
;             TELEFAX: 650/952-9881
;    INFORMATION FOR SEQ ID NO: 8:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 474 amino acids
;             TYPE: Amino Acid
;             TOPOLOGY: Linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-774-954-8

  Query Match             34.8%;  Score 123;  DB 11;  Length 474;
  Best Local Similarity   67.6%;  Pred. No. 6.1e-07;
  Matches   23;  Conservative    5;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         19 GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQH 52
              |||||| | ||| ||||| |:||| :| :||::|
Db        166 GRFGNQVDQFLGVLAFAKALDRTLVLPNFIEFKH 199


RESULT 9
US-10-820-474A-102
; Sequence 102, Application US/10820474A
; Publication No. US20050155089A1
; GENERAL INFORMATION:
;  APPLICANT: LAL, PREETI
;  APPLICANT:  TANG, Y. TOM
;  APPLICANT:  GORGONE, GINA A.
;  APPLICANT:  CORLEY, NEIL C.
;  APPLICANT:  GUEGLER, KARL J.
;  APPLICANT:  BAUGHN, MARIAH R.
;  APPLICANT:  AKERBLOM, INGRID E.
;  APPLICANT:  AU-YOUNG, JANICE
;  APPLICANT:  YUE, HENRY
;  APPLICANT:  PATTERSON, CHANDRA
;  APPLICANT:  REDDY, ROOPA
;  APPLICANT:  HILLMAN, JENNIFER L.
;  APPLICANT:  BANDMAN, OLGA
;  TITLE OF INVENTION: SIGNAL PEPTIDE-CONTAINING MOLECULES
;  FILE REFERENCE: 039386-1568
;  CURRENT APPLICATION NUMBER: US/10/820,474A
;  CURRENT FILING DATE:  2004-04-07
;  PRIOR APPLICATION NUMBER: 09/720,533
;  PRIOR FILING DATE: 2001-03-20
;  PRIOR APPLICATION NUMBER: PCT/US99/14484
;  PRIOR FILING DATE: 1999-06-25
;  PRIOR APPLICATION NUMBER: 60/090,762
;  PRIOR FILING DATE: 1998-06-26
;  PRIOR APPLICATION NUMBER: 60/094,983
;  PRIOR FILING DATE: 1998-07-31
;  PRIOR APPLICATION NUMBER: 60/102,686
;  PRIOR FILING DATE: 1998-10-01
;  NUMBER OF SEQ ID NOS: 269
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 102
;   LENGTH: 150
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte Clone No: 1647884
US-10-820-474A-102

  Query Match             34.0%;  Score 120;  DB 18;  Length 150;
  Best Local Similarity   95.0%;  Pred. No. 4.2e-07;
  Matches   19;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGR 20
              ||||||||||||||||||:
Db         24 MPAGSWDPAGYLLYCPCMGK 43
```

```
RESULT 10
US-09-774-954-14
; Sequence 14, Application US/09774954
; Publication No. US20040241645A1
;  GENERAL INFORMATION:
;       APPLICANT: Yang Wang, Michael W. Spellman
;       TITLE OF INVENTION: O-Fucosyltransferase
;       NUMBER OF SEQUENCES: 17
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Genentech, Inc.
;            STREET: 1 DNA Way
;            CITY: South San Francisco
;            STATE: California
;            COUNTRY: USA
;            ZIP: 94080
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: WinPatin (Genentech)
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/774,954
;            FILING DATE: 30-Jan-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/978,741
;            FILING DATE: 26-NOV-1997
;            APPLICATION NUMBER: 08/792,498
;            FILING DATE: 31-JAN-1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Svoboda, Craig G.
;            REGISTRATION NUMBER: 39,044
;            REFERENCE/DOCKET NUMBER: P1041P1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 650/225-1489
;            TELEFAX: 650/952-9881
;  INFORMATION FOR SEQ ID NO: 14:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 28 amino acids
;            TYPE: Amino Acid
;            TOPOLOGY: Linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-774-954-14

  Query Match             28.6%;  Score 101;  DB 11;  Length 28;
  Best Local Similarity   90.0%;  Pred. No. 1.8e-05;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGR 20
              |||||||||||||||| | |||
Db          9 MPAGSWDPAGYLLYXPXMGR 28


RESULT 11
US-10-425-115-327031
; Sequence 327031, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 327031
;   LENGTH: 110
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(110)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_61321C.1.pep
US-10-425-115-327031

  Query Match             24.4%;  Score 86;  DB 16;  Length 110;
  Best Local Similarity   47.8%;  Pred. No. 0.0067;
  Matches   22;  Conservative    2;  Mismatches   20;  Indels    2;  Gaps    2;

Qy          2 PAGSWDPA-GYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPP 46
              |   |||| ||:| |  |   ||   |  :   || ||||| | |
Db         58 PKSRWDPAQKYLVYLPFEG-ISNQFYSFQNAATMAKRLNRTLVVXP 102


RESULT 12
US-10-424-599-266118
; Sequence 266118, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT:  Kovalic David K
;  APPLICANT:  Zhou Yihua
;  APPLICANT:  Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 266118
;   LENGTH: 152
;   TYPE: PRT
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(152)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_82325C.1.pep
US-10-424-599-266118

  Query Match             22.8%;  Score 80.5;  DB 15;  Length 152;
  Best Local Similarity   41.9%;  Pred. No. 0.049;
  Matches   18;  Conservative    7;  Mismatches   15;  Indels    3;  Gaps    2;

Qy         11 YLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHH 53
              :: | |  | | ||   |  ::  | :||||| ||| ::  ||
Db         76 FVWYAPHSG-FSNQLSEFKNAVLMAGILNRTLVVPPILD--HH 115


RESULT 13
US-10-425-115-201329
; Sequence 201329, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(53222)B
;  CURRENT APPLICATION NUMBER: US/10/425,115
;  CURRENT FILING DATE:  2003-04-28
;  NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 201329
;   LENGTH: 189
;   TYPE: PRT
;   ORGANISM: Zea mays
```

```
;   FEATURE:
;   OTHER INFORMATION: Clone ID: MRT4577_115198C.1.pep
US-10-425-115-201329

  Query Match             22.0%;  Score 77.5;  DB 16;  Length 189;
  Best Local Similarity   45.5%;  Pred. No. 0.15;
  Matches   20;  Conservative    3;  Mismatches   18;  Indels    3;  Gaps    2;

Qy         10 GYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHH 53
              | | | |  | | ||      :   | |||||| ||| ::  ||
Db         78 GGLWYAPHSG-FSNQVGELRNAAVAAALLNRTLVVPPVLD--HH 118


RESULT 14
US-10-437-963-113919
; Sequence 113919, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 113919
;   LENGTH: 538
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_17661C.1.pep
US-10-437-963-113919

  Query Match             22.0%;  Score 77.5;  DB 16;  Length 538;
  Best Local Similarity   44.2%;  Pred. No. 0.47;
  Matches   19;  Conservative    4;  Mismatches   17;  Indels    3;  Gaps    2;

Qy         11 YLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHH 53
              :| | |  | | ||      :   | |||||| ||| ::  ||
Db        110 FLWYAPHSG-FSNQVGELRNAAVAAALLNRTLVVPPVLD--HH 149


RESULT 15
US-10-156-761-12551
; Sequence 12551, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12551
;   LENGTH: 601
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-12551

  Query Match             20.8%;  Score 73.5;  DB 14;  Length 601;
  Best Local Similarity   33.8%;  Pred. No. 1.7;
  Matches   26;  Conservative    9;  Mismatches   21;  Indels   21;  Gaps    6;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGS---LAFAKL----LN--RTLAVPP----- 46
              :|  |||   | |: | :| :| |  | ||:   || | :    ||   || |||
Db         56 LPKNSWDIKNY-LWAPRLGMYGIQRIHLLGNVMVLAGAGVGGGSLNYANTLYVPPKAFFD 114

Qy         47 ---W---IEYQHHKPPF 57
                 |    ::|    |:
Db        115 DPQWKDITDWQEELKPY 131


Search completed: October 25, 2005, 15:41:23
Job time : 35.8234 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:25:36 ; Search time 10.0205 Seconds
(without alignments)
454.426 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	353	100.0	61	3	US-08-978-741-9	Sequence 9, Appli
2	353	100.0	61	3	US-09-333-729A-4	Sequence 4, Appli
3	353	100.0	365	3	US-08-978-741-2	Sequence 2, Appli
4	353	100.0	365	3	US-09-333-729A-3	Sequence 3, Appli
5	353	100.0	397	3	US-08-978-741-6	Sequence 6, Appli
6	353	100.0	397	3	US-09-333-729A-7	Sequence 7, Appli
7	295	83.6	61	3	US-08-978-741-3	Sequence 3, Appli
8	295	83.6	61	3	US-09-333-729A-5	Sequence 5, Appli
9	223	63.2	399	4	US-09-270-767-45921	Sequence 45921, A
10	218	61.8	343	3	US-08-978-741-17	Sequence 17, Appl
11	218	61.8	343	3	US-09-333-729A-13	Sequence 13, Appl
12	123	34.8	474	3	US-08-978-741-8	Sequence 8, Appli
13	123	34.8	474	3	US-09-333-729A-12	Sequence 12, Appl
14	101	28.6	28	3	US-08-978-741-14	Sequence 14, Appl
15	101	28.6	28	3	US-09-333-729A-14	Sequence 14, Appl
16	63.5	18.0	205	4	US-09-540-236-3055	Sequence 3055, Ap
17	62.5	17.7	340	2	US-08-578-592-5	Sequence 5, Appli
18	62.5	17.7	340	3	US-09-185-111-5	Sequence 5, Appli
19	62	17.6	412	4	US-09-640-211A-2258	Sequence 2258, Ap
20	60.5	17.1	737	4	US-09-583-110-5075	Sequence 5075, Ap
21	60.5	17.1	763	4	US-09-107-433-3610	Sequence 3610, Ap
22	60.5	17.1	783	2	US-08-922-837-2	Sequence 2, Appli
23	60.5	17.1	783	3	US-09-351-550-2	Sequence 2, Appli
24	59.5	16.9	398	4	US-09-830-428A-5	Sequence 5, Appli
25	59	16.7	786	4	US-09-543-681A-6379	Sequence 6379, Ap
26	58.5	16.6	315	4	US-09-902-540-16137	Sequence 16137, A
27	58	16.4	333	5	PCT-US91-00899-11	Sequence 11, Appl
28	58	16.4	365	1	US-07-914-281-6	Sequence 6, Appli
29	58	16.4	365	1	US-08-393-246-6	Sequence 6, Appli
30	58	16.4	365	1	US-08-273-411-1	Sequence 1, Appli
31	58	16.4	365	1	US-08-525-058A-6	Sequence 6, Appli
32	58	16.4	365	1	US-08-395-800A-6	Sequence 6, Appli
33	58	16.4	365	1	US-08-395-800A-10	Sequence 10, Appl
34	58	16.4	365	2	US-08-696-731-6	Sequence 6, Appli
35	58	16.4	365	3	US-09-042-531-6	Sequence 6, Appli
36	58	16.4	365	3	US-09-151-592-2	Sequence 2, Appli
37	58	16.4	365	3	US-09-254-077A-10	Sequence 10, Appl
38	58	16.4	365	3	US-09-254-077A-11	Sequence 11, Appl
39	58	16.4	365	4	US-09-443-766-13	Sequence 13, Appl
40	58	16.4	365	5	PCT-US91-00899-12	Sequence 12, Appl
41	58	16.4	399	4	US-09-949-016-11353	Sequence 11353, A
42	57.5	16.3	461	3	US-09-457-046B-64	Sequence 64, Appl
43	57.5	16.3	461	4	US-09-866-570B-64	Sequence 64, Appl
44	56.5	16.0	288	4	US-09-252-991A-25627	Sequence 25627, A
45	56.5	16.0	350	4	US-09-489-039A-9626	Sequence 9626, Ap

ALIGNMENTS

RESULT 1
US-08-978-741-9
; Sequence 9, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
; APPLICANT: Yang Wang, Michael W. Spellman
; TITLE OF INVENTION: O-Fucosyltransferase
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,741
; FILING DATE: 26-No. 6100076-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/792498
; FILING DATE: 31
; ATTORNEY/AGENT INFORMATION:
; NAME: Svoboda, Craig G.
; REGISTRATION NUMBER: 39,044
; REFERENCE/DOCKET NUMBER: P1041P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-1489
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 61 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
US-08-978-741-9

Query Match 100.0%; Score 353; DB 3; Length 61;
Best Local Similarity 100.0%; Pred. No. 2.3e-40;
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
```

```
                |
Db          61 H 61


RESULT 2
US-09-333-729A-4
; Sequence 4, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 4
;   LENGTH: 61
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-333-729A-4

  Query Match             100.0%;  Score 353;  DB 3;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-40;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 3
US-08-978-741-2
; Sequence 2, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  365 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-2

  Query Match             100.0%;  Score 353;  DB 3;  Length 365;
  Best Local Similarity   100.0%;  Pred. No. 2e-39;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 4
US-09-333-729A-3
; Sequence 3, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 3
;   LENGTH: 365
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-333-729A-3

  Query Match             100.0%;  Score 353;  DB 3;  Length 365;
  Best Local Similarity   100.0%;  Pred. No. 2e-39;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 5
US-08-978-741-6
; Sequence 6, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
```

```
;       FILING DATE:  26-No. 6100076-1997
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/792498
;       FILING DATE:  31
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Svoboda, Craig G.
;       REGISTRATION NUMBER:  39,044
;       REFERENCE/DOCKET NUMBER:  P1041P1
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650/225-1489
;       TELEFAX:  650/952-9881
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  397 amino acids
;       TYPE:  Amino Acid
;       TOPOLOGY:  Linear
US-08-978-741-6

  Query Match             100.0%;  Score 353;  DB 3;  Length 397;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-39;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         33 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 92

Qy         61 H 61
              |
Db         93 H 93


RESULT 6
US-09-333-729A-7
; Sequence 7, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 7
;   LENGTH: 397
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid insert encoded protein.
US-09-333-729A-7

  Query Match             100.0%;  Score 353;  DB 3;  Length 397;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-39;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         33 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 92

Qy         61 H 61
              |
Db         93 H 93


RESULT 7
US-08-978-741-3
; Sequence 3, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  61 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-3

  Query Match             83.6%;  Score 295;  DB 3;  Length 61;
  Best Local Similarity   91.5%;  Pred. No. 1.6e-32;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||
Db          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 8
US-09-333-729A-5
; Sequence 5, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 5
;   LENGTH: 61
;   TYPE: PRT
;   ORGANISM: Cricetulus Grieseus
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 15, 17, 38
;   OTHER INFORMATION: unknown amino acid
US-09-333-729A-5

  Query Match             83.6%;  Score 295;  DB 3;  Length 61;
  Best Local Similarity   91.5%;  Pred. No. 1.6e-32;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
```

```
                  ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||
Db          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61


RESULT 9
US-09-270-767-45921
; Sequence 45921, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45921
;   LENGTH: 399
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-270-767-45921

  Query Match             63.2%;  Score 223;  DB 4;  Length 399;
  Best Local Similarity   84.4%;  Pred. No. 8.6e-22;
  Matches   38;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          7 DPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQ 51
              || ||| |||||||||||||||||||||| ||||| :|||:||:
Db         25 DPNGYLTYCPCMGRFGNQADHFLGSLAFAKALNRTLILPPWVEYR 69


RESULT 10
US-08-978-741-17
; Sequence 17, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  343 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-17

  Query Match             61.8%;  Score 218;  DB 3;  Length 343;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-21;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         23 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              |||||||||||||||||||||||||||||||||||||||
Db          1 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 39


RESULT 11
US-09-333-729A-13
; Sequence 13, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 13
;   LENGTH: 343
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-09-333-729A-13

  Query Match             61.8%;  Score 218;  DB 3;  Length 343;
  Best Local Similarity   100.0%;  Pred. No. 3.4e-21;
  Matches   39;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         23 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              |||||||||||||||||||||||||||||||||||||||
Db          1 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 39


RESULT 12
US-08-978-741-8
; Sequence 8, Application US/08978741
; Patent No. 6100076
;  GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  474 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-8

  Query Match             34.8%;  Score 123;  DB 3;  Length 474;
  Best Local Similarity   67.6%;  Pred. No. 3.6e-08;
  Matches   23;  Conservative    5;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         19 GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQH 52
              |||||| | ||| ||||| |:||| :| :||::|
Db        166 GRFGNQVDQFLGVLAFAKALDRTLVLPNFIEFKH 199


RESULT 13
US-09-333-729A-12
; Sequence 12, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 12
;   LENGTH: 474
;   TYPE: PRT
;   ORGANISM: Caenorhabditis Elegans
US-09-333-729A-12

  Query Match             34.8%;  Score 123;  DB 3;  Length 474;
  Best Local Similarity   67.6%;  Pred. No. 3.6e-08;
  Matches   23;  Conservative    5;  Mismatches    6;  Indels    0;  Gaps    0;

Qy         19 GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQH 52
              |||||| | ||| ||||| |:||| :| :||::|
Db        166 GRFGNQVDQFLGVLAFAKALDRTLVLPNFIEFKH 199


RESULT 14
US-08-978-741-14
; Sequence 14, Application US/08978741
; Patent No. 6100076
; GENERAL INFORMATION:
;    APPLICANT:  Yang Wang, Michael W. Spellman
;    TITLE OF INVENTION:  O-Fucosyltransferase
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,741
;      FILING DATE:  26-No. 6100076-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/792498
;      FILING DATE:  31
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Svoboda, Craig G.
;      REGISTRATION NUMBER:  39,044
;      REFERENCE/DOCKET NUMBER:  P1041P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-1489
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  28 amino acids
;      TYPE:  Amino Acid
;      TOPOLOGY:  Linear
US-08-978-741-14

  Query Match             28.6%;  Score 101;  DB 3;  Length 28;
  Best Local Similarity   90.0%;  Pred. No. 1.1e-06;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGR 20
              |||||||||||||||| | |||
Db          9 MPAGSWDPAGYLLYXPXMGR 28


RESULT 15
US-09-333-729A-14
; Sequence 14, Application US/09333729A
; Patent No. 6270987
; GENERAL INFORMATION:
;  APPLICANT:  Wang,Yang
;  APPLICANT:  Spellman,Michael W.
;  TITLE OF INVENTION:  O-Fucosyltransferase
;  FILE REFERENCE:  P1041P1D1-Substitute
;  CURRENT APPLICATION NUMBER: US/09/333,729A
;  CURRENT FILING DATE:  1999-06-15
;  PRIOR APPLICATION NUMBER:  US 08/798,741
;  PRIOR FILING DATE:  1997-11-26
;  NUMBER OF SEQ ID NOS:  21
; SEQ ID NO 14
;   LENGTH: 28
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid insert encoded protein.
;   NAME/KEY: unsure
;   LOCATION: 23, 25
;   OTHER INFORMATION: unknown amino acid
US-09-333-729A-14

  Query Match             28.6%;  Score 101;  DB 3;  Length 28;
  Best Local Similarity   90.0%;  Pred. No. 1.1e-06;
  Matches   18;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGR 20
              |||||||||||||||| | |||
Db          9 MPAGSWDPAGYLLYXPXMGR 28


Search completed: October 25, 2005, 15:42:49
Job time : 11.0205 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: October 25, 2005, 15:20:23 ; Search time 38.8296 Seconds
(without alignments)
607.588 Million cell updates/sec

Title: US-09-774-954-9
Perfect score: 353
Sequence: 1 MPAGSWDPAGYLLYCPCMGR.....LAVPPWIEYQHHKPPFTNLH 61

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	353	100.0	61	2	AAW80574	Aaw80574 N-termina
2	353	100.0	365	2	AAW80571	Aaw80571 Human hea
3	353	100.0	388	8	ABM82223	Abm82223 Tumour-as
4	353	100.0	397	2	AAW80573	Aaw80573 Human O-f
5	295	83.6	61	2	AAW80572	Aaw80572 N-termina
6	241.5	68.4	417	4	AAM78377	Aam78377 Human pro
7	229.5	65.0	417	4	AAM79361	Aam79361 Human pro
8	223	63.2	402	4	ABB63790	Abb63790 Drosophil
9	218	61.8	343	2	AAW80577	Aaw80577 Partial h
10	120	34.0	150	3	AAY87325	Aay87325 Human sig
11	101	28.6	28	2	AAW80575	Aaw80575 N-termina
12	69.5	19.7	1561	5	ABB97437	Abb97437 Novel hum
13	64.5	18.3	223	6	ABU35781	Abu35781 Protein e
14	64	18.1	366	6	ABM69163	Abm69163 Photorhab
15	63.5	18.0	191	6	ABU34983	Abu34983 Protein e
16	63.5	18.0	205	8	ADL05369	Adl05369 M. catarr
17	63.5	18.0	580	8	ADN25627	Adn25627 Bacterial
18	62.5	17.7	340	2	AAR63066	Aar63066 Acetyl-es
19	62	17.6	412	3	AAB33303	Aab33303 Pinus rad
20	60.5	17.1	737	8	ADK48560	Adk48560 Streptoco
21	60.5	17.1	741	6	ABU01274	Abu01274 S. pneumo
22	60.5	17.1	763	8	ADR94975	Adr94975 Novel S.
23	60.5	17.1	767	4	AAU37750	Aau37750 Streptoco
24	60.5	17.1	767	4	AAU38005	Aau38005 Streptoco
25	60.5	17.1	767	6	ABU45987	Abu45987 Protein e
26	60.5	17.1	768	3	AAY81596	Aay81596 Streptoco
27	60.5	17.1	783	2	AAW97416	Aaw97416 A membran
28	59.5	16.9	198	6	ABU36823	Abu36823 Protein e
29	59.5	16.9	198	6	ABU34438	Abu34438 Protein e
30	59.5	16.9	198	6	ABU33866	Abu33866 Protein e
31	59.5	16.9	398	3	AAY91094	Aay91094 Human G p
32	59.5	16.9	398	4	AAU08997	Aau08997 Human G p
33	59.5	16.9	398	4	AAE03544	Aae03544 Human G-p
34	59.5	16.9	398	4	AAM51415	Aam51415 Human OT7
35	59.5	16.9	398	5	ABB79864	Abb79864 Human G-p
36	59.5	16.9	398	6	ABP97222	Abp97222 Tumour-as
37	59.5	16.9	398	6	ABP81995	Abp81995 Human G p
38	59.5	16.9	398	7	ADN39505	Adn39505 Cancer/an
39	59.5	16.9	398	7	ADN39430	Adn39430 Cancer/an
40	59.5	16.9	398	8	ADG19755	Adg19755 Human G p
41	59.5	16.9	398	8	ADL91585	Adl91585 Human imm
42	59.5	16.9	398	8	ADO29412	Ado29412 Human GPC
43	59.5	16.9	398	8	ADQ89080	Adq89080 Human uro
44	59.5	16.9	398	8	ADR21325	Adr21325 Human met
45	59.5	16.9	398	8	ADT05941	Adt05941 Human OT7

ALIGNMENTS

RESULT 1
AAW80574
ID AAW80574 standard; peptide; 61 AA.
XX
AC AAW80574;
XX
DT 16-DEC-1998 (first entry)
XX
DE N-terminal amino acid sequence of human heart O-fucosyltransferase.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS Homo sapiens.
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Claim 4; Page 69; 90pp; English.
XX
CC This represents a the N-terminal sequence of the human heart O-
CC fucosyltransferase that can glycosylate an epidermal growth factor (EGF)
CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC domains, are used in diagnosis and treatment of conditions associated
CC with overexpression of O-fucosyltransferase, to promote survival of
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC used to detect gene amplification and expression. The expression can also
CC be determined at the protein level using antibodies specific for O-
CC fucosyltransferase
XX
SQ Sequence 61 AA;

```
  Query Match             100.0%;  Score 353;  DB 2;  Length 61;
  Best Local Similarity   100.0%;  Pred. No. 4.8e-40;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 2
AAW80571
ID   AAW80571 standard; protein; 365 AA.
XX
AC   AAW80571;
XX
DT   16-DEC-1998  (first entry)
XX
DE   Human heart O-fucosyltransferase.
XX
KW   O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW   O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1. .61
FT                   /note= "N-terminal sequence claimed for in claim 4"
XX
PN   WO9833924-A1.
XX
PD   06-AUG-1998.
XX
PF   17-DEC-1997;   97WO-US023401.
XX
PR   31-JAN-1997;   97US-00792498.
PR   26-NOV-1997;   97US-00978741.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wang Y,  Spellman MW;
XX
DR   WPI; 1998-437477/37.
DR   N-PSDB; AAV65632.
XX
PT   Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT   domains - useful for diagnosis and treatment of diseases involving
PT   overexpression of the enzyme.
XX
PS   Example 1; Fig 12A; 90pp; English.
XX
CC   This represents a human heart O-fucosyltransferase that can glycosylate
CC   an epidermal growth factor (EGF) domain of a polypeptide with an
CC   activated O-fucose residue. Inhibitors of O-fucosyltransferase, e.g.
CC   mutants with increased affinity for the EFG domains, are used in
CC   diagnosis and treatment of conditions associated with overexpression of O
CC   -fucosyltransferase, to promote survival of sensory (retinal) neurons.
CC   Probes based on EGF domain polypeptide are used to detect gene
CC   amplification and expression. The expression can also be determined at
CC   the protein level using antibodies specific for O-fucosyltransferase
XX
SQ   Sequence 365 AA;

  Query Match             100.0%;  Score 353;  DB 2;  Length 365;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-39;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60

Qy         61 H 61
              |
Db         61 H 61


RESULT 3
ABM82223
ID   ABM82223 standard; protein; 388 AA.
XX
AC   ABM82223;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Tumour-associated antigenic target (TAT) polypeptide PRO83146, SEQ:5726.
XX
KW   Tumour-associated antigenic target; TAT; human; overexpression; cancer;
KW   tumour; diagnosis; cell proliferative disorder; breast cancer;
KW   colorectal cancer; lung cancer; ovarian cancer; liver cancer;
KW   central nervous system cancer; bladder cancer; pancreatic cancer;
KW   cervical cancer; melanoma; leukaemia; hybridisation probe;
KW   chromosome identification; chromosome mapping; gene mapping;
KW   gene therapy; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004030615-A2.
XX
PD   15-APR-2004.
XX
PF   29-SEP-2003; 2003WO-US028547.
XX
PR   02-OCT-2002; 2002US-0414971P.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Wu TD,  Zhang Z,  Zhou Y;
XX
DR   WPI; 2004-347921/32.
DR   N-PSDB; ACN40746.
XX
PT   New tumor-associated antigenic target polypeptides and nucleic acids,
PT   useful in preparing a medicament for treating or detecting a
PT   proliferative disorder, e.g. breast, lung, colorectal, ovarian or
PT   prostate cancer or tumor.
XX
PS   Claim 12; SEQ ID NO 5726; 7273pp; English.
XX
CC   The invention relates to human tumour-associated antigenic target (TAT)
CC   polypeptides, and their related nucleic acids. The TAT polypeptides are
CC   overexpressed in cancer tissues compared to normal tissues, and may thus
CC   serve as effective targets for the diagnosis and treatment of cancer in
CC   mammals. The invention also relates to nucleic acid and polypeptide
CC   sequences at least 80% identical to the TAT nucleic acids and
CC   polypeptides; expression vectors and host cells comprising a TAT nucleic
CC   acid; an antibody specific for a TAT polypeptide; a peptide or organic
CC   molecule which binds to a TAT polypeptide; fusion proteins comprising a
CC   TAT polypeptide; and methods and compositions for the treatment or
CC   diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
CC   antibodies, antagonists, binding molecules and compositions are useful
CC   for diagnosing or treating a cell proliferative disorder associated with
CC   increased TAT expression, particularly cancers such as breast cancer,
CC   colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
CC   cancer, pancreatic cancer, cervical cancer, cancers of the central
CC   nervous system, melanoma and leukaemia. TAT nucleic acids may further be
CC   used as hybridisation probes, in chromosome and gene mapping, in
CC   chromosome identification and in gene therapy. The present sequence
CC   represents a TAT polypeptide of the invention
XX
SQ   Sequence 388 AA;

  Query Match             100.0%;  Score 353;  DB 8;  Length 388;
```

```
  Best Local Similarity   100.0%;  Pred. No. 4.1e-39;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 83

Qy         61 H 61
              |
Db         84 H 84
```

RESULT 4
AAW80573
ID AAW80573 standard; protein; 397 AA.
XX
AC AAW80573;
XX
DT 16-DEC-1998 (first entry)
XX
DE Human O-fucosyltransferase sequence expressed by a plasmid insertion.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key   Location/Qualifiers
FT Peptide   27. .32
FT   /note= "polyhistidine tag"
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
DR N-PSDB; AAV65633.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Example; Fig 13B; 90pp; English.
XX
CC This represents the amino acid sequence of the human heart O-
CC fucosyltransferase expressed by a plasmid insertion. The human O-
CC fucosyltransferase can glycosylate an epidermal growth factor (EGF)
CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC domains, are used in diagnosis and treatment of conditions associated
CC with overexpression of O-fucosyltransferase, to promote survival of
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC used to detect gene amplification and expression. The expression can also
CC be determined at the protein level using antibodies specific for O-
CC fucosyltransferase
XX
SQ Sequence 397 AA;

```
  Query Match             100.0%;  Score 353;  DB 2;  Length 397;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-39;
  Matches   61;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         33 MPAGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNL 92

Qy         61 H 61
              |
Db         93 H 93
```

RESULT 5
AAW80572
ID AAW80572 standard; peptide; 61 AA.
XX
AC AAW80572;
XX
DT 16-DEC-1998 (first entry)
XX
DE N-terminal amino acid sequence of CHO O-fucosyltransferase.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Misc-difference 15
FT   /note= "unknown"
FT Misc-difference 17
FT   /note= "unknown"
FT Misc-difference 38
FT   /note= "unknown"
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Claim 5; Page 69; 90pp; English.
XX
CC This represents the N-terminal sequence of CHO O-fucosyltransferase. The
CC enzyme can glycosylate an epidermal growth factor (EGF) domain of a
CC polypeptide with an activated O-fucose residue. Inhibitors of O-
CC fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC domains, are used in diagnosis and treatment of conditions associated
CC with overexpression of O-fucosyltransferase, to promote survival of
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC used to detect gene amplification and expression. The expression can also
CC be determined at the protein level using antibodies specific for O-
CC fucosyltransferase
XX
SQ Sequence 61 AA;

```
  Query Match             83.6%;  Score 295;  DB 2;  Length 61;
  Best Local Similarity   91.5%;  Pred. No. 3.8e-32;
  Matches   54;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          3 AGSWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              ||||| |||||| | ||||||||||||||||||||  ||||||||||||||||||||||
Db          3 AGSWDLAGYLLYXPXMGRFGNQADHFLGSLAFAKLXVRTLAVPPWIEYQHHKPPFTNLH 61
```

```
RESULT 6
AAM78377
ID   AAM78377 standard; protein; 417 AA.
XX
AC   AAM78377;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 1039.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y;
PI   Ma Y,  Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK51510.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 20; Page 3276-3277; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 417 AA;

  Query Match             68.4%;  Score 241.5;  DB 4;  Length 417;
  Best Local Similarity   76.7%;  Pred. No. 6.8e-24;
  Matches   46;  Conservative    3;  Mismatches    6;  Indels    5;  Gaps    1;

Qy          7 DPAGYLLYCPCM-----GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              | |  :  : | :     ||||||||||||||||||||||||||||||||||||||||||
Db         54 DMATWPTHAPSVEEGGQGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 113


RESULT 7
AAM79361
ID   AAM79361 standard; protein; 417 AA.
XX
AC   AAM79361;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 3007.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y;
PI   Ma Y,  Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK52494.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 20; Page 229-230; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 417 AA;

  Query Match             65.0%;  Score 229.5;  DB 4;  Length 417;
  Best Local Similarity   73.3%;  Pred. No. 2.9e-22;
  Matches   44;  Conservative    4;  Mismatches    7;  Indels    5;  Gaps    1;

Qy          7 DPAGYLLYCPCM-----GRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
              | |  :  : | :     ||||||||||||||||||||||:|||| ||||||||||||||
Db         54 DMATWPTHAPSVEEGGQGRFGNQADHFLGSLAFAKLLNRSLAVPSWIEYQHHKPPFTNLH 113


RESULT 8
ABB63790
ID   ABB63790 standard; protein; 402 AA.
```

XX
AC ABB63790;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 18162.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL07893.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 18162; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 402 AA;

Query Match 63.2%; Score 223; DB 4; Length 402;
Best Local Similarity 84.4%; Pred. No. 2.2e-21;
Matches 38; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 7 DPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPPWIEYQ 51
|| ||| |||||||||||||||||||||| ||||| :|||:||:
Db 28 DPNGYLTYCPCMGRFGNQADHFLGSLAFAKALNRTLILPPWVEYR 72

RESULT 9
AAW80577
ID AAW80577 standard; protein; 343 AA.
XX
AC AAW80577;
XX
DT 16-DEC-1998 (first entry)
XX
DE Partial human sequence of unknown function from a myeblast cell line.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS Homo sapiens.
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
DR N-PSDB; AAV65634.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Example 1; Page 62-68; 90pp; English.
XX
CC This represents a published partial human sequence of unknown function
CC from a myeblast cell line. The invention provides a human heart O-
CC fucosyltransferase enzyme that can glycosylate an epidermal growth factor
CC (EGF) domain of a polypeptide with an activated O-fucose residue.
CC Inhibitors of O-fucosyltransferase, e.g. mutants with increased affinity
CC for the EFG domains, are used in diagnosis and treatment of conditions
CC associated with overexpression of O-fucosyltransferase, to promote
CC survival of sensory (retinal) neurons. Probes based on EGF domain
CC polypeptide are used to detect gene amplification and expression. The
CC expression can also be determined at the protein level using antibodies
CC specific for O-fucosyltransferase
XX
SQ Sequence 343 AA;

Query Match 61.8%; Score 218; DB 2; Length 343;
Best Local Similarity 100.0%; Pred. No. 8.6e-21;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 23 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 61
|||||||||||||||||||||||||||||||||||||||
Db 1 NQADHFLGSLAFAKLLNRTLAVPPWIEYQHHKPPFTNLH 39

RESULT 10
AAY87325
ID AAY87325 standard; protein; 150 AA.
XX
AC AAY87325;
XX
DT 11-MAY-2000 (first entry)
XX
DE Human signal peptide containing protein HSPP-102 SEQ ID NO:102.
XX
KW Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
KW inflammation; cardiovascular disease; anticancer; anti-inflammatory;
KW antimicrobial; nootropic; neuroprotective; cardiovascular; hepatotropic;
KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
KW reproductive disorder; developmental disorder; arteriosclerosis;
KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
KW Parkinson's disease; Huntington's diseases; ovulatory defect;
KW muscular dystrophy.
XX
OS Homo sapiens.
XX
PN WO200000610-A2.
XX
PD 06-JAN-2000.
XX
PF 25-JUN-1999; 99WO-US014484.
XX
PR 26-JUN-1998; 98US-0090762P.
PR 31-JUL-1998; 98US-0094983P.

PR 01-OCT-1998; 98US-0102686P.
PR 11-DEC-1998; 98US-0112129P.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
PI Bandman O;
XX
DR WPI; 2000-160673/14.
DR N-PSDB; AAZ98210.
XX
PT New human signal peptide-containing proteins useful in treatment,
PT prevention and diagnosis of e.g. cancer, inflammation and cardiovascular
PT disease.
XX
PS Claim 1; Page 225-226; 327pp; English.
XX
CC AAZ98109 to AAZ98242 encode AAY87224 to AAY87357 which represent the
CC human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
CC anticancer, anti-inflammatory, antimicrobial, nootropic, hepatotropic,
CC neuroprotective, cardiovascular and antiasthmatic activities, and can be
CC used in gene therapy. HSPPs can be used to treat or prevent disorders
CC associated with decreased activity or function of HSPP. Antagonists of
CC HSPP are used to treat or prevent disorders associated with increased
CC activity or function of HSPP. Such diseases include cell proliferation
CC (including cancer), inflammation, cardiovascular, neurological,
CC reproductive or developmental disorders, (e.g. arteriosclerosis,
CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
CC asthma, Crohn's disease, microbial or other infections, congestive or
CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
CC nucleic acids can be used for the recombinant production of HSPP, for
CC detecting HSPP in standard hybridisation and amplification assays (for
CC diagnosis and monitoring), in gene therapy, as antisense, triplex-forming
CC or ribozyme therapeutics, for detecting related sequences or genetic
CC variations, and for chromosomal mapping. HSPP are also used to raise
CC specific antibodies (Ab) and to screen for agonists and antagonists
CC (potential therapeutic agents). Ab are used to diagnose, or monitor, HSPP
CC -related diseases (in usual immunoassays), as therapeutic antagonists, in
CC competitive drug screens, and for purification of HSPP from natural
CC sources
XX
SQ Sequence 150 AA;

Query Match 34.0%; Score 120; DB 3; Length 150;
Best Local Similarity 95.0%; Pred. No. 7.3e-08;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPAGSWDPAGYLLYCPCMGR 20
||||||||||||||||||:
Db 24 MPAGSWDPAGYLLYCPCMGK 43

RESULT 11
AAW80575
ID AAW80575 standard; protein; 28 AA.
XX
AC AAW80575;
XX
DT 16-DEC-1998 (first entry)
XX
DE N-terminal sequence of expressed human O-fucosyltransferase.
XX
KW O-fucosyltransferase; epidermal growth factor; EGF; glycosylation;
KW O-fucose; inhibitor; sensory neuron; retinal neuron; human; heart.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 23
FT /note= "unknown"
FT Misc-difference 25
FT /note= "unknown"
XX
PN WO9833924-A1.
XX
PD 06-AUG-1998.
XX
PF 17-DEC-1997; 97WO-US023401.
XX
PR 31-JAN-1997; 97US-00792498.
PR 26-NOV-1997; 97US-00978741.
XX
PA (GETH) GENENTECH INC.
XX
PI Wang Y, Spellman MW;
XX
DR WPI; 1998-437477/37.
XX
PT Human O-fucosyl:transferase able to glycosylate epidermal growth factor
PT domains - useful for diagnosis and treatment of diseases involving
PT overexpression of the enzyme.
XX
PS Example 1; Page 39; 90pp; English.
XX
CC This represents the N-terminal amino acid sequence of the human heart O-
CC fucosyltransferase expressed by a plasmid insertion. The human O-
CC fucosyltransferase can glycosylate an epidermal growth factor (EGF)
CC domain of a polypeptide with an activated O-fucose residue. Inhibitors of
CC O-fucosyltransferase, e.g. mutants with increased affinity for the EFG
CC domains, are used in diagnosis and treatment of conditions associated
CC with overexpression of O-fucosyltransferase, to promote survival of
CC sensory (retinal) neurons. Probes based on EGF domain polypeptide are
CC used to detect gene amplification and expression. The expression can also
CC be determined at the protein level using antibodies specific for O-
CC fucosyltransferase
XX
SQ Sequence 28 AA;

Query Match 28.6%; Score 101; DB 2; Length 28;
Best Local Similarity 90.0%; Pred. No. 4e-06;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MPAGSWDPAGYLLYCPCMGR 20
|||||||||||||| | |||
Db 9 MPAGSWDPAGYLLYXPXMGR 28

RESULT 12
ABB97437
ID ABB97437 standard; protein; 1561 AA.
XX
AC ABB97437;
XX
DT 27-JUN-2002 (first entry)
XX
DE Novel human protein SEQ ID NO: 705.
XX
KW Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag.
XX
OS Homo sapiens.
XX
PN WO200222660-A2.
XX
PD 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US026015.
XX
PR 11-SEP-2000; 2000US-00659671.
XX
PA (HYSE-) HYSEQ INC.

XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-292408/33.
DR N-PSDB; ABN32623.
XX
PT An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX
PS Claim 20; SEQ ID NO 705; 509pp; English.
XX
CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention
XX
SQ Sequence 1561 AA;

Query Match 19.7%; Score 69.5; DB 5; Length 1561;
Best Local Similarity 29.1%; Pred. No. 8.4;
Matches 16; Conservative 7; Mismatches 23; Indels 9; Gaps 2;

Qy 5 SWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAVPP---WIEYQHHKPP 56
Db 1342 TWEQGRYLIWNPCSGHFYGQFD------TFCPLKNVGCLIGPDNIWFNIQRYESP 1390

RESULT 13
ABU35781
ID ABU35781 standard; protein; 223 AA.
XX
AC ABU35781;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #21308.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Mycobacterium leprae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA39651.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 63705; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 223 AA;

Query Match 18.3%; Score 64.5; DB 6; Length 223;
Best Local Similarity 43.8%; Pred. No. 4.2;
Matches 14; Conservative 4; Mismatches 11; Indels 3; Gaps 1;

Qy 21 FGNQADHFLGSLAFAKLLNRTLAV---PPWIE 49
Db 141 FGRQVDSFEGDIGFAGLVDPVRAVFIRAPWVE 172

RESULT 14
ABM69163
ID ABM69163 standard; protein; 366 AA.
XX
AC ABM69163;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens protein sequence #2260.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
OS Photorhabdus luminescens.
XX
PN WO200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002WO-IB003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;

```
PI   Buchrieser C;
XX
DR   WPI; 2003-148459/14.
XX
PT   Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT   useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS   Claim 2; SEQ ID NO 2260; 1205pp; French.
XX
CC   The invention relates to the isolation of genes and their encoded
CC   proteins from Photorhabdus luminescens. The isolated sequences are
CC   sources of probes and primers for detecting the genome of P. luminescens
CC   and related species; to study polymorphisms; for gene analysis and for
CC   detection/amplification of the genes. Antibodies (Ab) raised against the
CC   polypeptides encoded by the genes are used for detection/identification
CC   of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC   carry a gene-containing vector are used to select compounds that
CC   modulate, regulate, induce or inhibit expression of the genes in plants,
CC   animals or microorganisms other than P. luminescens and are able to alter
CC   response or sensitivity to toxins and antibiotics produced by P.
CC   luminescens. Cells transformed to express the genes are useful for
CC   recombinant production of the proteins, particularly toxins and
CC   antibacterials useful as insecticides, bactericides and fungicides. The
CC   genes, proteins, vectors containing the genes and Ab are also useful
CC   therapeutically (to treat microbial infection by bacteria or fungi that
CC   are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC   biopesticides. Other uses of the genes and the proteins are as virulence
CC   factors and for identifying targets of human diseases for which P.
CC   luminescens is a model (particularly plague and whooping cough). This
CC   sequence represents one of the isolated P. luminescens proteins
XX
SQ   Sequence 366 AA;

  Query Match             18.1%;  Score 64;  DB 6;  Length 366;
  Best Local Similarity   33.3%;  Pred. No. 8.7;
  Matches   21;  Conservative    6;  Mismatches   16;  Indels   20;  Gaps    3;

Qy          5 SWDPAGYLLYCPCMGRFGNQADHFLGSLAFAKLLNRTLAV----------------PPWI 48
              :|| | | ||   || | |  |: | || |  :|   :|:                |||
Db        163 TWDGAAYTLY--LMGTFVN--DYGLYSLIFLVILFILVALSLPFLGGRLRTGLDRFPPWS 218

Qy         49 EYQ 51
                |:
Db        219 LYR 221


RESULT 15
ABU34983
ID   ABU34983 standard; protein; 191 AA.
XX
AC   ABU34983;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #20510.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Moraxella catarrhalis.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA38853.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 62907; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 191 AA;

  Query Match             18.0%;  Score 63.5;  DB 6;  Length 191;
  Best Local Similarity   38.8%;  Pred. No. 4.8;
  Matches   19;  Conservative    4;  Mismatches   13;  Indels   13;  Gaps    3;

Qy         25 ADH-------FLGSLAFAKLLNRTLAVPPWI----EYQHHK--PPFTNL 60
              |:|       |  :|  |:||  |  |  |    |||:||  |||  |
Db         68 AEHEAASCGEILDEVAIKKILNMPLRAPMIIVIATEYQYHKKVPPFEQL 116


Search completed: October 25, 2005, 15:30:42
Job time : 40.8296 secs
```